

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 17:34:09 ; Search time 4339 Seconds

(without alignments)
11028.050 Million cell updates/sec

Title: US-10-041-615-33

Sequence: 1 atgataaactagccctta.....gtgtcatctgtcttatgact 1104

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hvg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hvg_hum:*
31: em_hvg_inv:*
32: em_hvg_other:*
33: em_hvg_mus:*
34: em_hvg_pln:*
35: em_hvg_rnd:*
36: em_hvg_mam:*
37: em_hvg_vrt:*
38: em_sy:*
39: em_hvg_hum:*
40: em_hvg_mus:*
41: em_hvg_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Match	Length	DB	ID	Description
1	1104	100.0	1104	6	AX702273	AX702273 Sequence
2	779	70.6	4319	6	AX647101	AX647101 Sequence
3	779	70.6	98861	6	AL136961	AL136961 Human DNA
4	551	49.9	151174	2	AC016189	AC016189 Homo sapi
5	397.2	36.0	536	6	AX230148	AX230148 Sequence
6	378.4	34.3	556	6	AX244735	AX244735 Sequence
7	353.4	32.0	151174	2	AC016189	AC016189 Homo sapi
8	261.2	23.7	162543	2	EX470183	EX470183 Danto rer
9	221.6	20.1	190588	2	AC102718	AC102718 Mus muscu
10	221.6	20.1	204438	2	AC110249	AC110249 Mus muscu
11	220	19.9	1493	10	BC052868	BC052868 Mus muscu
12	208.6	18.9	239344	2	AC133310	AC133310 Rattus no
13	208.6	18.9	244625	2	AC094739	AC094739 Rattus no
14	184.8	16.7	1086	6	AR308611	AR308611 Sequence
15	184.8	16.7	1086	6	AX704977	AX704977 Sequence
16	184.8	16.7	1643	6	AR003350	AR003350 Sequence
17	184.8	16.7	1643	6	AR269382	AR269382 Sequence
18	184.8	16.7	1643	6	AR270716	AR270716 Sequence
19	184.8	16.7	1643	6	AX774793	AX774793 Sequence
20	184.8	16.7	1643	9	HMGPCRB	LOC1177 Human EBV 1
21	183.2	16.6	1086	6	AR308705	AR308705 Sequence
22	183.2	16.6	1638	6	AX548826	AX548826 Sequence
23	183.2	16.6	1710	9	BC020752	BC020752 Homo sapi
24	183.2	16.6	149478	9	AL160155	AL160155 Human DNA
25	165	14.9	187201	5	EX004885	EX004885 Zebrafish
26	119.6	10.8	1238	4	CCU33334	U33334 Oryctolagus
27	119.4	10.8	1020	6	AX458308	AX458308 Sequence
28	119.4	10.8	1020	6	BD187417	BD187417 Novel pol
29	119.4	10.8	1076	6	AX375230	AX375230 Sequence
30	118.6	10.7	2025	6	AR270819	AR270819 Sequence
31	118.6	10.7	2025	6	AX548931	AX548931 Sequence
32	118.6	10.7	2025	9	HSU07225	U07225 Homo sapien
33	118.2	10.7	1382	4	AB051422	AB051422 Sus scrofa
34	117.8	10.7	184377	10	AC121579	AC121579 Mus muscu
35	117.8	10.7	195290	2	AC135635	AC135635 Mus muscu
36	117	10.6	1134	9	AY136753	AY136753 Homo sapi
37	117	10.6	2257	9	BC028135	BC028135 Homo sapi
38	117	10.6	2681	9	BC012104	BC012104 Homo sapi
39	116.4	10.6	123185	2	AP000587	AP000587 Homo sapi
40	116.4	10.5	196988	9	AP002761	AP002761 Homo sapi
41	113.4	10.3	250610	2	AC120742	AC120742 Rattus no
42	113.4	10.3	276127	2	AC116265	AC116265 Rattus no
43	112	10.1	1020	6	AR308577	AR308577 Sequence
44	112	10.1	1020	6	AR308693	AR308693 Sequence
45	112	10.1	1741	9	HS294154	Z94154 H.sapiens m

ALIGNMENTS

RESULT 1
AX702273
LOCUS AX702273 1104 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 33 from Patent WO02064793.
ACCESSION AX702273
VERSION AX702273.1 GI:29537547
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 Casman,S.J., Edinger,S.R., Ellerman,K., Smithson,G., Kekuda,R. and
Muralidhara,P.
TITLE Novel spcr-like proteins and nucleic acids encoding same

JOURNAL Patent: WO 02064793-A 33 22-AUG-2002;

FEATURES Curagen Corporation (US)
Location/Qualifiers

1..1104
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
1..1083
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD87117.1"
/db_xref="GI:29537548"
/db_xref="EMBL:CAD87117"
/translation="MILGPFETPTKIKIKIKDLNVSQTVKIFLRSLEYTALVLS
ALGNIALCTOKSRKINCTGIYLVHVAVDLFTVALPGRVVCYVIGSSWPGKGL
CRLEFAVLYTDYGVYLMACVSDHYPAVAVCAWGPRLRTGRARVCAVATWLVLL
OTMPLLMPTKPLVYKLCMEYSMSVGLPLMVIVAFIAGCGVGIILSCYMKI
TWKLCSTARPDPTVRKGRHWRGCLITLMVAVVCPSPYHNIKQPMARGLHLS
CAERAPFLSDATYALNMNNGITPIPIYFPASTHYRKMILGILKMGSSSSSSSS
TPGKASSETPSITQARGSMFLAEHYV"

ORIGIN

Query Match 100.0%; Score 1104; DB 6; Length 1104;
Best Local Similarity 100.0%; Pred. No. 2,3e-215;
Matches 1104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGATPAACTAGGCCCTTACTTCAACACCAACCAAAATTAATTAAGT 60
1 ATGATPAACTAGGCCCTTACTTCAACACCAACCAAAATTAATTAAGT 60
61 TTAATGAGAAAGCCAACTGGAATTTTTCGCTCTCTTCAACAGCCCTC 120
61 TTAATGAGAAAGCCAACTGGAATTTTTCGCTCTCTTCAACAGCCCTC 120
121 CTGCTCTAGTGCCTGGGAAACATCTTGCCTTTGCTTACTGTAAGACGCG 180
121 CTGCTCTAGTGCCTGGGAAACATCTTGCCTTTGCTTACTGTAAGACGCG 180
181 AAGATPAACTAGGCCCTTACTTCAACACCAACCAAAATTAATTAAGT 240
181 AAGATPAACTAGGCCCTTACTTCAACACCAACCAAAATTAATTAAGT 240
241 GTGCTCTAGTGCCTGGGAAACATCTTGCCTTTGCTTACTGTAAGACGCG 300
241 GTGCTCTAGTGCCTGGGAAACATCTTGCCTTTGCTTACTGTAAGACGCG 300
301 GGGCTCTAGTGCCTGGGAAACATCTTGCCTTTGCTTACTGTAAGACGCG 360
301 GGGCTCTAGTGCCTGGGAAACATCTTGCCTTTGCTTACTGTAAGACGCG 360
361 ATGCTCTAGTGCCTGGGAAACATCTTGCCTTTGCTTACTGTAAGACGCG 420
361 ATGCTCTAGTGCCTGGGAAACATCTTGCCTTTGCTTACTGTAAGACGCG 420
421 CTGCTCTAGTGCCTGGGAAACATCTTGCCTTTGCTTACTGTAAGACGCG 480
421 CTGCTCTAGTGCCTGGGAAACATCTTGCCTTTGCTTACTGTAAGACGCG 480
481 CACAGATGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 540
481 CACAGATGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 540
541 ATGAGATGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 600
541 ATGAGATGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 600
601 GCATGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 660
601 GCATGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 660
661 AAGCTGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 720
661 AAGCTGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 720

721 GGCTGCTCTAGTGCCTGGGAAACATCTTGCCTTTGCTTACTGTAAGACGCG 780
721 GGCTGCTCTAGTGCCTGGGAAACATCTTGCCTTTGCTTACTGTAAGACGCG 780
781 CTCAATATGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 840
781 CTCAATATGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 840
841 AAGCTGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 900
841 AAGCTGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 900
901 ACCCTGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 960
901 ACCCTGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 960
961 AAGCTGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 1020
961 AAGCTGAGCCCTTCTGATGCCATGACCAACGCGCTGAGTGGGCAAGCTGCG 1020
1021 TCAGAAACACCAAGTATCAACCAAGCCAGAGGTTCTATGTTCTAGCGAGCAGTGTTC 1080
1021 TCAGAAACACCAAGTATCAACCAAGCCAGAGGTTCTATGTTCTAGCGAGCAGTGTTC 1080
1081 TAAGCTGATGATGCTTATGACT 1104
1081 TAAGCTGATGATGCTTATGACT 1104

RESULT 2
AX647101 4319 bp DNA linear PAT 04-MAR-2003
LOCUS
DEFINITION
Sequence 1293 from Patent EP1270724.
ACCESSION
AX647101
VERSION
AX647101.1 GI:28800143
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 Suwa, M., Asai, K., Akiyama, Y. and Abuzaitani, H.
Guanosine triphosphate-binding protein coupled receptors
TITLE
Patent: EP 1270724-A 1293 02-JAN-2003;
JOURNAL
National Institute of Advanced Industrial Science and Technology
(JP) ; Center for Advanced Science and Technology Incubation, Ltd.
(JP)

FEATURES
Source
1..4319
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
join(201..308,3128..3531,3660..4119)
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD69676.1"
/db_xref="GI:28800144"
/translation="MFVDSVGDIEBGTGMIFLGLTEFVNAEKDQGLAFROGAL
QADGVCAHLHRLHGVYLMACVSDHYPAVAVCAWGPRLRTGRARVCAVATWLVLLQ
TMPLLMPTKPLVYKLCMEYSMSVGLPLMVIVAFIAGCGVGIILSCYMKI
WKLCSTARPDPTVRKGRHWRGCLITLMVAVVCPSPYHNIKQPMARGLHLS
LOPLPPOHVAHERAPPAIILRAGSLFVPSGHRGQHEHLMHYPIIFVFASTHYR
KMLGITLKGSSSSSSSSSTGKASSETPSITQARGSMFLAEHYV"

ORIGIN

Query Match 70.6%; Score 779; DB 6; Length 4319;
Best Local Similarity 83.6%; Pred. No. 6,9e-149;
Matches 1006; Conservative 0; Mismatches 0; Indels 197; Gaps 3;

```

Db      2940 CGCTCTCTGTCTACACAGCCCTCTGCTTCACTGTCCTGAGAAACATCTTGGCTT 2999
Qy      157 TGCCCTACCTGTCAAAAAGACAGAAAGATCACTGACAGGACTTACTGCTGACCTG 216
Db      3000 TGCCCTACCTGTCAAAAAGACAGAAAGATCACTGACAGGACTTACTGCTGACCTG 3059
Qy      217 GCTGTGTCTGACCTGCTGCTTCACTGCTGCTTACCGGAAAGGTGTGTATGTGCTG 276
Db      3060 GCTGTGTCTGACCTGCTGCTTCACTGCTGCTTACCGGAAAGGTGTGTATGTGCTG 3119
Qy      277 GGCCTCAGCTGAGCTTTCGCGCAAGGAGCTTGCAGGCTGACCGGCTTTGTGCTTACAC 336
Db      3120 GGCCTCAGCTGAGCTTTCGCGCAAGGAGCTTGCAGGCTGACCGGCTTTGTGCTTACAC 3179
Qy      337 GACACCTACAGGAGGAGGCTTACCTCACTGCTGCTGTGTGAGGCTGACCATTAACCACTG 396
Db      3180 GACACCTACAGGAGGAGGCTTACCTCACTGCTGCTGTGTGAGGCTGACCATTAACCACTG 3238
Qy      397 GTCGTGCTCAGCTGAGGAGGAGGCTTCCGCAAGGAGCTGCGCGCAGGCTGCTGCTG 456
Db      3239 GTCGTGCTCAGCTGAGGAGGAGGCTTCCGCAAGGAGCTGCGCGCAGGCTGCTGCTG 3298
Qy      457 GCCATCTGACCTTGTGCTGCTGACAGAGCTGCTGCTTGTATGCTTATGCTTACCAAG 516
Db      3299 GCCATCTGACCTTGTGCTGCTGACAGAGCTGCTGCTTGTATGCTTATGCTTACCAAG 3358
Qy      517 CCGGTGTGTGAGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 576
Db      3359 CCGGTGTGTGAGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3418
Qy      577 CCGCTCATGTGTGCTGCTGCTTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 636
Db      3419 CCGCTCATGTGTGCTGCTGCTTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 3478
Qy      637 TCTGTGTATATGAAGATCACTTGAAGCTGTGACGACACT----- 678
Db      3479 TCTGTGTATATGAAGATCACTTGAAGCTGTGACGACACTGAGAAACCAAGTACCA 3538
Qy      679 ----- 678
Db      3539 GCGGAGAAAGGACACACCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3598
Qy      679 ----- 678
Db      3599 CGCCACCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3658
Qy      679 -----CGG 681
Db      3659 GGAACAGCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3718
Qy      682 GAGAGCCCAAGTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 741
Db      3719 GAGAGCCCAAGTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3778
Qy      742 ATGTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 801
Db      3779 ATGTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3838
Qy      802 GCGAGAGGAGTGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 861
Db      3839 GCGAGAGGAGTGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3898
Qy      862 CAGGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 921
Db      3899 CAGGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3957
Qy      922 GCATCCACCCATTAAGAGAAATGCTCTGCGCATTTTAAAGCTCAAAAGGCTCTTCTCC 981
Db      3958 GCATCCACCCATTAAGAGAAATGCTCTGCGCATTTTAAAGCTCAAAAGGCTCTTCTCC 4017
Qy      982 TCTCTCTCTCTCTCTCTCTCTCAACCCCAAGAAAGGCTTCTTCAAAAACCAAGATAC 1041
Db      4018 TCTCTCTCTCTCTCTCTCTCTCAACCCCAAGAAAGGCTTCTTCAAAAACCAAGATAC 4077

```

```

Qy      1042 CAGGCGCAGGAGTGTATGTTCTTACGCGAGCAGTGTCTTACGCTGCTTATG 1101
Db      4078 CAGGCGCAGGAGTGTATGTTCTTACGCGAGCAGTGTCTTACGCTGCTTATG 4137
Qy      1102 ACT 1104
Db      4138 ACT 4140

RESULT 3
AL136961/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Apr 12, 2001 this sequence version replaced gi:13776997.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HeP/Chr13
RP11-178C10 is from the library RP11-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-178C10 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP11-178C10 is at 98861 in this
sequence. The true right end of clone RP11-461N23 is at 100 in this
sequence.

FEATURES
source
location/Qualifiers
1..98861
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-178C10"
/clone_1ib="RP11-11.1"
161..463
/clone="RP11M8 repeat: matches 5836..6160 of consensus"
566..723
/clone="RP11M8 repeat: matches 2592..2659 of consensus"
repeat_region
repeat_region

```


Qy	1042	CAGGCAGAGGTTCTATGTTCTTACGGAGCAGCGTGTCTAACGTGTCATTGCTTTATG	1101
Db	55784	CAGGCAGAGGTTCTATGTTCTTACGGAGCAGCGTGCTAACGTGTCATTGCTTTATG	55725
Qy	1102	ACT 1104	
Db	55724	ACT 55722	

RESIT.T 4

AC016189
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE

SOURCE
ORGANIZATION

REFERENCE
AUTHOR
TITLE
JOURNAL
REFERENCE
AUTHOR

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	Excellent article, well-researched and informative.
2. The Impact of Technology on Learning	Journal of Educational Technology	Very good, provides a comprehensive overview of the topic.
3. The Importance of Social Skills in Education	Journal of Social and Clinical Psychology	Good article, but could have included more recent research.
4. The Effect of Parental Involvement on Student Achievement	Journal of Family Psychology	Excellent, well-written and easy to read.
5. The Role of the School in Promoting Student Health	Journal of School Health	Good article, but needs more data to support the conclusions.
6. The Impact of Teacher Stress on Student Learning	Journal of Educational Psychology	Very good, provides a detailed analysis of the issue.
7. The Role of the Teacher in Promoting Student Creativity	Journal of Creative Behavior	Excellent, well-researched and informative.
8. The Impact of Teacher Expectations on Student Achievement	Journal of Educational Research	Good article, but needs more data to support the conclusions.
9. The Role of the Teacher in Promoting Student Self-Esteem	Journal of Educational Psychology	Excellent, well-written and easy to read.
10. The Impact of Teacher Burnout on Student Learning	Journal of Educational Psychology	Very good, provides a detailed analysis of the issue.

Direct Submission
Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:16939915.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://www.genome.washington.edu/PW/RepeatMasker.html>

```

* NOTE: This record contains 152 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

*	1	851:	contig of 851 bp in length
*			are preserved.
*	852	951:	gap of 100 bp
*	952	1804:	contig of 853 bp in length
*	1805	1904:	gap of 100 bp
*	1905	2719:	contig of 815 bp in length
*	2720	2819:	gap of 100 bp

* 2820 3689: contig of 849 bp in length
* 3689: gap of 100 bp
* 3789: contig of 889 bp in length
* 4657: gap of 100 bp
* 4658: contig of 843 bp in length
* 4758: gap of 100 bp
* 5601: contig of 832 bp in length
* 5701: gap of 100 bp
* 6533: gap of 100 bp
* 6633: contig of 849 bp in length
* 7481: gap of 100 bp
* 7482: gap of 100 bp
* 7582: contig of 866 bp in length
* 8447: gap of 100 bp
* 8448: contig of 836 bp in length
* 8548: gap of 100 bp
* 9384: gap of 100 bp
* 9484: gap of 100 bp
* 10343: gap of 100 bp
* 10443: contig of 881 bp in length
* 11323: gap of 100 bp
* 11424: contig of 831 bp in length
* 12254: gap of 100 bp
* 12354: contig of 934 bp in length
* 13288: gap of 100 bp
* 13289: gap of 100 bp
* 13389: contig of 847 bp in length
* 14236: gap of 100 bp
* 14336: contig of 837 bp in length
* 15173: gap of 100 bp
* 15273: contig of 858 bp in length
* 16130: gap of 100 bp
* 16131: contig of 843 bp in length
* 16231: gap of 100 bp
* 17073: contig of 830 bp in length
* 17173: gap of 100 bp
* 18003: contig of 836 bp in length
* 18104: gap of 100 bp
* 18939: contig of 846 bp in length
* 19040: gap of 100 bp
* 19886: contig of 849 bp in length
* 19986: gap of 100 bp
* 20835: gap of 100 bp
* 20935: contig of 827 bp in length
* 21762: gap of 100 bp
* 21763: contig of 882 bp in length
* 22744: gap of 100 bp
* 22844: contig of 865 bp in length
* 23709: gap of 100 bp
* 23809: contig of 868 bp in length
* 24677: gap of 100 bp
* 24777: contig of 826 bp in length
* 25602: gap of 100 bp
* 25702: contig of 858 bp in length
* 26560: gap of 100 bp
* 26660: contig of 844 bp in length
* 27504: gap of 100 bp
* 27604: contig of 845 bp in length
* 28443: gap of 100 bp
* 28543: contig of 849 bp in length
* 29399: gap of 100 bp
* 29499: contig of 859 bp in length
* 30358: gap of 100 bp
* 31323: contig of 866 bp in length
* 31423: gap of 100 bp
* 32300: contig of 877 bp in length
* 32400: gap of 100 bp
* 33254: contig of 854 bp in length
* 33255: gap of 100 bp
* 33355: contig of 864 bp in length
* 34218: gap of 100 bp
* 34318: contig of 851 bp in length
* 35169: gap of 100 bp
* 35269: contig of 856 bp in length
* 36123: gap of 100 bp
* 36226: contig of 836 bp in length
* 37061: gap of 100 bp
* 37161: contig of 842 bp in length
* 38003: gap of 100 bp

* 38004 38103: gap of 100 bp
* 38104: contig of 842 bp in length
* 38945: gap of 100 bp
* 39045: contig of 837 bp in length
* 39882: gap of 100 bp
* 39883: contig of 851 bp in length
* 40833: gap of 100 bp
* 40933: contig of 846 bp in length
* 41779: gap of 100 bp
* 41879: contig of 843 bp in length
* 42722: gap of 100 bp
* 42822: contig of 856 bp in length
* 43678: gap of 100 bp
* 43778: contig of 856 bp in length
* 44634: gap of 100 bp
* 44734: contig of 853 bp in length
* 45587: gap of 100 bp
* 45687: contig of 855 bp in length
* 46542: gap of 100 bp
* 46642: contig of 846 bp in length
* 47488: gap of 100 bp
* 47588: contig of 860 bp in length
* 48448: gap of 100 bp
* 48548: contig of 902 bp in length
* 49450: gap of 100 bp
* 49550: contig of 833 bp in length
* 50383: gap of 100 bp
* 50483: contig of 870 bp in length
* 51353: gap of 100 bp
* 51453: contig of 853 bp in length
* 52306: gap of 100 bp
* 52406: contig of 846 bp in length
* 53252: gap of 100 bp
* 53352: contig of 872 bp in length
* 54224: gap of 100 bp
* 54324: contig of 838 bp in length
* 55162: gap of 100 bp
* 55262: contig of 824 bp in length
* 56086: gap of 100 bp
* 56186: contig of 840 bp in length
* 57026: gap of 100 bp
* 57126: contig of 899 bp in length
* 58025: gap of 100 bp
* 58125: contig of 879 bp in length
* 59004: gap of 100 bp
* 59104: contig of 875 bp in length
* 59779: gap of 100 bp
* 60079: contig of 841 bp in length
* 60920: gap of 100 bp
* 61020: contig of 865 bp in length
* 61885: gap of 100 bp
* 61985: contig of 807 bp in length
* 62792: gap of 100 bp
* 62892: contig of 852 bp in length
* 63744: gap of 100 bp
* 63844: contig of 849 bp in length
* 64693: gap of 100 bp
* 64793: contig of 852 bp in length
* 65645: gap of 100 bp
* 65745: contig of 879 bp in length
* 66624: gap of 100 bp
* 66724: contig of 826 bp in length
* 67550: gap of 100 bp
* 67650: contig of 852 bp in length
* 68502: gap of 100 bp
* 68602: contig of 833 bp in length
* 69435: gap of 100 bp

Query Match 49.9%; Score 551; DB 2; Length 151174;

Best Local Similarity 97.8%; Pred. No. 3.2e-102;

Matches 590; Conservative 0; Mismatches 10; Indels 3; Gaps 3;

Qy 126 CTTACGTCCCTGGGAACATCCTTGCCTTGTACCTGTCA-AAAGAGCGGAGA 184

```

Query Match      36.0%; Score 397.2; DB 6; Length 536;
Best Local Similarity 97.5%; Pred. No. 7.7e-71;
Matches 435; Conservative 0; Mismatches 8; Indels 3; Gaps 3;

ORIGIN

      285  CTGGCCTTTTCGGCAAGGGGCTCTGCAGGCTGACGGGGTTTGCTCTCACCGCACCTA 344
           CTGGCCTTTTCGGCAAGGGGCTCTGCAGGCTGACGGGGTTTGCTCTCACCGCACCTA 67

```

[illegible]**Db**

```

QY 937 AGGAATGGCTCTGGGCAATTTTAAAGCTCAAGGGTCTTCTCTCTCTCTCTCTCTCTCC 996
Db 317 AGGAATGGCTCTGGGCAATTTTAAAGCTCAAGGGTCTTCTCTCTCTCTCTCTCTCC 261
QY 997 TCTCCACCCAGGAAAGCTCTTTCAGAAACACCAAGTATACCCAGGCGAGGTTCT 1056
Db 260 TCTCCACCCAGGAAAGCTCTTTCAGAAACACCAAGTATACCCAGGCGAGGTTCT 201
QY 1057 ATCTTCTTAGCGAGCACGTGGTCTAACGTGTCTATTCGTTTATGACT 1104
Db 200 ATCTTCTTAGCGAGCACGTGGTCTAACGTGTCTATTCGTTTATGACT 153

RESULT 7
AC016189 151174 bp DNA linear HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-23M9, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC016189
ACCESSION AC016189.3 GI:9129134
VERSION HTG; HTGS_PHASE0.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 151174)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens, clone RP11-23M9
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 151174)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckler, R., Boguslavsky, L., Bouckgalter, B.,
Brown, A., Castle, A., Collangelo, M., Collins, S., Collamore, A.,
Cooke, P., DeArillano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,
Leshocky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meidrim, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6939915.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4399
Center clone name: 23_M_9
-----
* NOTE: This record contains 152 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 851: contig of 851 bp in length
* 951: gap of 100 bp
* 952 1804: contig of 853 bp in length

```

```

1805 1904: gap of 100 bp
1905 2019: contig of 815 bp in length
2720 2819: gap of 100 bp
2820 3668: contig of 849 bp in length
3669 3768: gap of 100 bp
3769 4657: contig of 889 bp in length
4658 4757: gap of 100 bp
4758 5600: contig of 843 bp in length
5601 5700: gap of 100 bp
5701 5832: contig of 832 bp in length
5833 6632: gap of 100 bp
6633 7481: contig of 849 bp in length
7482 7581: gap of 100 bp
7582 8447: contig of 866 bp in length
8448 8547: gap of 100 bp
8548 9383: contig of 836 bp in length
9384 9483: gap of 100 bp
9484 10342: contig of 859 bp in length
10343 10442: gap of 100 bp
10443 11223: contig of 881 bp in length
11224 11423: gap of 100 bp
11424 12254: contig of 831 bp in length
12255 12354: gap of 100 bp
12355 13288: contig of 934 bp in length
13289 14235: gap of 100 bp
14236 14335: contig of 847 bp in length
14336 15172: contig of 837 bp in length
15173 15272: gap of 100 bp
15273 16130: contig of 858 bp in length
16131 16230: gap of 100 bp
16231 17073: contig of 843 bp in length
17074 17173: gap of 100 bp
17174 18003: contig of 830 bp in length
18004 18103: gap of 100 bp
18104 18939: contig of 836 bp in length
18940 19039: gap of 100 bp
19040 19885: contig of 846 bp in length
19886 19985: gap of 100 bp
19986 20834: contig of 849 bp in length
20835 20934: gap of 100 bp
20935 21761: contig of 827 bp in length
21762 21861: gap of 100 bp
21862 22743: contig of 882 bp in length
22744 22843: gap of 100 bp
22844 23708: contig of 865 bp in length
23709 23808: gap of 100 bp
23809 24676: contig of 868 bp in length
24677 24776: gap of 100 bp
24777 25602: contig of 826 bp in length
25603 25702: gap of 100 bp
25703 25860: contig of 858 bp in length
25861 26660: gap of 100 bp
26661 27504: contig of 844 bp in length
27505 27604: gap of 100 bp
27605 28449: contig of 845 bp in length
28450 28549: gap of 100 bp
28550 29398: contig of 849 bp in length
29399 29498: gap of 100 bp
29499 30357: contig of 859 bp in length
30358 30457: gap of 100 bp
30458 31233: contig of 866 bp in length
31234 31423: gap of 100 bp
31424 32300: contig of 877 bp in length
32301 32400: gap of 100 bp
32401 33254: contig of 854 bp in length
33255 33354: gap of 100 bp
33355 34218: contig of 864 bp in length
34219 34318: gap of 100 bp
34319 35169: contig of 851 bp in length
35170 35269: gap of 100 bp
35270 36125: contig of 856 bp in length
36126 36225: gap of 100 bp

```


* consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 5005: contig of 5005 bp in length
 * 5106 5105: gap of 100 bp
 * 5106 101540: contig of 96435 bp in length
 * 101541 101540: gap of 100 bp
 * 101541 156970: contig of 55330 bp in length
 * 101541 156971 157070: gap of 100 bp
 * 156971 162543: contig of 5473 bp in length.
 * 157071

FEATURES

Location/Qualifiers
 1..162543
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="CH211-184W13"
 /clone_lib="CHORI-211"

misc_feature

1..5005
 /note="assembly fragment:01046
 fragment chain:1
 clone_end:SP6
 vector_side:left"

misc_feature

5106..101540
 /note="assembly fragment:01887
 fragment chain:1
 101641..156970
 /note="assembly fragment:00490
 fragment chain:2
 clone_end:17
 vector_side:right"

misc_feature

157071..162543
 /note="assembly fragment:01214
 fragment chain:2"

misc_feature

ORIGIN

Query Match 23.7%; Score 261.2; DB 2; Length 162543;
 Best Local Similarity 56.6%; Pred. No. 5.8e-43;
 Matches 504; Conservative 0; Mismatches 383; Indels 3; Gaps 1;
 QY 101 CTCTGTTCTACACAGCCCTCTCGTCTTCAGTGCCTCGGAAACATCTTCCCTTTGCC 160
 DB 133490 CACTTTTCTACTGCTGATTTTGTGATCAGGTGCGAGTAACCTCCCTGCTCTGCA 133431
 QY 161 TTACCTGTCAAAGAGCAGGAAGATCACTGCACAGGCATCTACCTGTGCACCTGGCTG 220
 DB 133430 TCACCTGTGAGAAAGCGCAGAGATGAATCTTACCACAAATATTTGATCAACTGGCCA 133371
 QY 221 TGTCTGACCTGTGTTTCACCGTGGCTTACCGGAGGGTGTGTGTTATGTCTGGGCT 280
 DB 133370 TCTCCGATGCTCTCTTCACTTAGTCTCTGCCGCGCAGGATCACTTACTACATCCGTGGCT 133311
 QY 281 CCAGCTGGCTTTGGCAAGGGCTCTGACGCTACGCGCTTGTGCTCTACACCGACA 340
 DB 133310 TTGATTGGCTTTTGAGATTTCTTGTGCGGCTGACGCTATGATCTTCTACAGCAATA 133251
 QY 341 CCTACGGGGGGTCTACTCATGGCTGTGTGAGGTGGACCATACCCAGCTGTGGTCT 400
 DB 133250 CCTATGATGATCGGTTTCTATGACCTGATCAGTGTGATCGTACTTGGCCATGCTGC 133191
 QY 401 GTGCCCACTGGGCGCGCTCCGACAGGCTGGCGCGCCAGGCTGTGCTCGTGGCCA 460
 DB 133190 ACCGTGAGGCTCTCAAGACTGAGAAAAACAAAGTCTGAAAGCGCTCTGTGTTCTGG 133131
 QY 461 TCTGGACCTTGGTCTGCTGCAGACCATGCTTGTGATGCCCATGACCAAGCCGC 520
 DB 133130 TGTGGTGGTGGTCTCTTTGAAACATCACTATGCTTTTACAGCATCATGGAACA 133071
 QY 521 TGTGGGCAAGCTGGCTGCTGATGAGTACAGCAGCATGAGTCACTCGGCTGCCCC 580

DB 133070 GACCTGGTCAATCGAAGCTGCATGGAGTTCTCCATTTTGACAG---CAAGAGGATGGCCT 133014
 QY 581 TCATGCTGCTGCTGGTGGCCCTTTGCCATTTGCTGTGGGCCAGTGGGGATCATCTGTCTCT 640
 DB 133013 ATGTTCTTCTGTTGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 132954
 QY 641 GCTATATGAGATCACTTGGAGCTGTGCAGCACAGCTCGGAGGACCCAGCTGACCAACA 700
 DB 132953 CCTACAGCGGGTCACTGCAAGCTCTCAAGACGGCCAAAGAAACCTTTGACCAACA 132894
 QY 701 GGAAGAGAGCCACTGGCGAGGCTGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 760
 DB 132893 AGTCTGGAGGCCACACACAGAGCCAAATATGATCTTACTTCTGCTGCTGCTGCTGCTG 132834
 QY 761 TGTGCTTACGCCCCCTACACCTCAACATCAAGCAGTCTCATGCGAGAGGATGCTCCACC 820
 DB 132833 TGTGCTTACGCCCCCTACACATTAACATCATGAGTTTGTGTGAGAGGCTTTTACCAG 132774
 QY 821 TGCATCTCTGTCGCGAGGAGGCTTTCTTACTGCTCTTCCCTTACGCCACCGTGGCCCTCA 880
 DB 132773 AGCCCTCTTGTGAAGAGTTTAAAGGCTTTGAAATTTGCTCTGCAGGTAACCGTTGCCATGA 132714
 QY 881 TGAACATGAGCTGTGGCATTACCCCAATCACTTCTTCTGATCCATCCATCCATTCACAGGA 940
 DB 132713 TGAACATTAACAGCTGCTTGAACCTTATTTACTTTTTTCCCATGACAGCATTAAGC 132654
 QY 941 AATGGCTCTGCGGCAATTTTAAAGCTCAAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 990
 DB 132653 AGAGTCAATGAGTCTTTTAAAGAGTACATTTCCATTTCCGAGCAGCTCC 132604

RESULT 9

AC102718 190588 bp DNA linear HTG 14-JUN-2003
 LOCUS Mus musculus clone RP24-469E6, WORKING DRAFT SEQUENCE, 9 unordered
 DEFINITION pieces:
 AC102718 GI:31745314
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Birren, B., Nusbaum, C. and Lander, E.
 Mus musculus, clone RP24-469E6
 Unpublished
 2 (bases 1 to 190588)

REFERENCE

1 (bases 1 to 190588)
 AUTHORS Birren, B., Nusbaum, C. and Lander, E.
 TITLE Mus musculus, clone RP24-469E6
 JOURNAL Unpublished
 2 (bases 1 to 190588)

REFERENCE

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B.,
 Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
 Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Katat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
 MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,
 Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigliio, J., Vassiliev, H.,
 Videl, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE

JOURNAL

```

REFERENCE 3 (bases 1 to 190588)
AUTHORS Birren,B., Nusbaum,C., Lander,B., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukagaiter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArelilano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
-----
Direct Submission
Submitted (14-JUN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 14, 2003 this sequence version replaced gi:22381711.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L19343
Center clone name: 469 E 6
-----
Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 187732 bases at least Q40
Consensus quality: 18884 bases at least Q30
Consensus quality: 189429 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 189788; sum-of-contigs
Quality coverage: 10.2 in Q20 bases; agarose-fp
Quality coverage: 9.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 330: contig of 330 bp in length
* 331 430: gap of 100 bp
* 431 970: contig of 540 bp in length
* 971 1070: gap of 100 bp
* 1071 7799: contig of 6729 bp in length
* 7800 7899: gap of 100 bp
* 7900 16142: contig of 8243 bp in length
* 16143 16243: gap of 100 bp
* 16243 29388: contig of 13146 bp in length
* 29389 29488: gap of 100 bp
* 29489 43346: contig of 19858 bp in length
* 43347 49347: gap of 100 bp
* 49347 101875: contig of 52429 bp in length
* 101876 101976: gap of 100 bp
* 101976 136840: contig of 34865 bp in length
* 136841 136941: gap of 100 bp
* 136941 190588: contig of 53648 bp in length.
*

```

```

FEATURES
source
Location/Qualifiers
1..190588
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone_lib="RPCI-24 Male Mouse BAC"
misc_feature
1..330
/notes="assembly_fragment"
clone_end:SP6
vector_side:left
misc_feature
431..970
/notes="assembly_fragment"
misc_feature
1071..7799
/notes="assembly_fragment"
misc_feature
7900..16142
/notes="assembly_fragment"
misc_feature
16243..29388
/notes="assembly_fragment"
misc_feature
29489..49346
/notes="assembly_fragment"
misc_feature
49447..101875
/notes="assembly_fragment"
misc_feature
101976..136840
/notes="assembly_fragment"
misc_feature
136941..190588
/notes="assembly_fragment"
clone_end:rf7
vector_side:right
ORIGIN
Query Match 20.1%; Score 221.6; DB 2; Length 190588;
Best Local Similarity 53.1%; Pred. No. 7.3e-35;
Matches 498; Conservative 0; Mismatches 434; Indels 6; Gaps 1;
QY 101 CTCGTGTTACACAGCCCTCTGCTTTCAGTGGCCCTGGGAAACATCTTGCCTTTGCC 160
DB 107260 CTCGCAATTACAGCCTGCTTTCATCATTTGGGCTGGTGGGAAACCTCTGCGCTTGTG 107319
QY 161 TTACTGTCAAAGAGCAGGAGGATCAACTGCACAGGATCTACTGCTGGTGCACCTGGCTG 220
DB 107320 TCATTGTTCAAACAGAAAAAATCAACTCAACACATCTTATTCATGAACCTGTGCA 107379
QY 221 TGTCTGACCTCTGTTTACCCGTGCCCTTACCGGAAAGGGTGTGTATGTGTGTGGCT 280
DB 107380 TTTCTGACATCTGTTTACCACAGCTTACCACATCGGATAGCCCTACTGCGCTGGCT 107439
QY 281 CCAGCTGSCCTTTCCGCAAGGGGCTTCGAGGCTGACGGGTTTGTGCTCTACACCGACA 340
DB 107440 TTGATTGGAGGATAGGTGATGCCCTGTGCGGGTAACCTGCTCTGCTGTTCATCAACA 107499
QY 341 CTTACGGGGGGTCTTACCTCATGGCCTGTGTGAGCGTGGACCAATTACCCAGCTGTGCT 400
DB 107500 CGTACGCGAGGTGTGAACCTTCATGACTTGTGTCATAGACCGCTTCTTCGCTGTG 107559
QY 401 GTGCCCATTTGGGCGCGGCTTCGACGCTGCGCGGCGCCAGCTGGTCTGCGCTGGCGCA 460
DB 107560 ACCCTCTGCGCTTACAAAGAAATTAAGAAATCGAATACGAAAGGTTCTGCTCTGCT 107619
QY 461 TCTGGACCTTGTGTGCTGTGTCAGACGATGCCCTTGTCTTGTATGCCCATGACCAAGCCG 520
DB 107620 TCTGGATTCTGCTTTCCTCTCAACACCTGCGCTGCTCTCACCCCTATGCTAAGGAGG 107679
QY 521 TGGTGGCAAGCTGGCCCTGCATGAGAGTACAGAGATGGAGTCAGTCTCGGGCTGCCCC 580
DB 107680 AGGGAGACAAGACCACTTGCATGGAGTATCCAAACTTTTGAAGGGGACAGCGTCCCT 107739
QY 581 TCATGCTCTGCTGGCCCTTTGCAATTTGCTTCTGTGGGCCAGTGGGGATCATCTCTGCT 640
DB 107740 GGATTCTGCTCGAGCCCTGTCTGCTGGGCTACGCTGCTGCTATCACAGTCAATCTCT 107799
QY 641 GCTATATGAAGATCACCTGGAGCTGTGCAGCACAGCTCGGAGGAGCCAGCTGACACGCA 700

```

```

TITLE
JOURNAL
COMMENT

```


160314 TTGATTGGAGGATGAGTGATGCGCTGTGCGGGTAACTGCTGTGGTGTCTACATCAACA 160373
QY 341 CCTACGGGGGGTCTACCTCATGGCTGTGTGAGCTGTGAGCTGACCATACCCAGCTGTGGTCT 400
D 160374 CGTACGAGTGTGAATTCATGACTTGTGTTGAGCATAGACCGCTCTTCGCTGTGGTGC 160433
QY 401 GTGCCACATGGGGCCCGCGCTCCGACGGCTGTGGCGCGCCAGGCTGTGTCTCGTGGGCA 460
D 160434 ACCCTCTGGGTACAAAGATTAAGAAATCGAATACGCAAAAGGTGTGTCTGCTGTGG 160493
QY 461 TCTGGACCTTGGTCTGCTGCAGCATGCTTGTCTTGTATGCGCCATGACCAAGCCGC 520
D 160494 TCTGGATCTTGGTCTTGTCTCAAACTGCTGCGCTGTCTCTCACTATGCTCTAAGAGG 160553
QY 521 TGTGGGCAAGCTGGCTGCTGCATGGATGACAGCAGTGTGAGTCACTTCCTGGGCTGCC 580
D 160554 AGGAGACAGACCACTTGCATGGATGATCAAACTTTGAAGGAGCAGCGTCCCTGCC 160613
QY 581 TCATGCTCTGGTGGCTTGTGCAATGGCTTGTGTGGCCAGTGGGGATCATCTGTCTCT 640
D 160614 GGATCTGCTGGAGCTGTCTGCTGGGCTACGTGTGCTGTGCTATCAGTATCTCTCTGT 160673
QY 641 GCTATATGAGATCACCTGGAAGCTGTGCAGCAGCTCGGAGGACCCAGTGAACGCA 700
D 160674 GTTACTCTCAGATCTGCTGCACAACTCTTCAGGACTGCCAAGCAGAACCCACTCAC 160733
QY 701 GGAAGACGCCACTGCGAGGCTGCTGTGCTGAGTGTGCTGTGGTGGCTGTGGTGG 760
D 160734 AATCTGCTGTGAACAAAGAGCTCTCAACAAATATCTCTCATCTGTGCTGTCTATCC 160793
QY 761 TCTGCTTCAGCCCTACCACTCAACATCAAGCAGTTCATGGCGAGAGGATGCTCCACC 820
D 160794 TGTGCTTCAGCCCTACCACTGCGCATCATTCAGCAGATGATAAGATGCTCTGTCTCC 160853
QY 821 TGCATCTC-----TGTGCGGAGGAGGCTTCTTACTGTCTTCTGAGCCACCGTGG 874
D 160854 CTGGAGCCCTGGAGTGTGGGCGGAGACATTCCTTCCAGATCTCTCTGCACTTCA 160913
QY 875 CCCTCATGACATGAACTGTGGCATTCACCAATCATTTACTTCTTGTGATCCACCCAT 934
D 160914 GCTGTGTAATCTCACTGCTGATGAGACCGCTTCAATACCTTCTTGTGATGCAAGGT 160973
QY 935 ACAGGAATGGCTCTGGGCAATTTAAAGTCAAGGCTTCTCTCTCTCTCTCTCTCTCT 994
D 160974 ATAAGAGAAAGTTCATGAAGTGTCAAGCTCAAGTGTGCTGATCTCCAGGCGAG 161033
QY 995 CTTCTTCAACCCAGGAAAGCTTCTTCAGAAACCA 1032
D 161034 TGAGGTGAGCCCTGGAAGAAATTCGCGGAAATGACA 161071

RESULT 11

BC052868
LOCUS BC052868 1493 bp mRNA linear ROD 08-OCT-2003
DEFINITION Mus musculus Epstein-Barr virus induced gene 2, mRNA (cdna clone
MGC:60769 IMAGE:30073287), complete cds.
ACCESSION BC052868
VERSION BC052868.1 GI:31127264
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
1 (bases 1 to 1493)
REFERENCE
AUTHORS Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Hsieh, P.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.D.,
Scheetz, T.E., Brownstein, M.J., Udell, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,

Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., R.A.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallos, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

23388257

12477932

2 (bases 1 to 1493)

Straussberg, R.

Direct Submission

Submitted (27-MAY-2003)

National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: gcaps-remail.nih.gov

Tissue Procurement: Drs. Dennis Taub, Dan Longo (NIA, USA),

Jonathan Keller (NCI, USA)

cDNA Library Preparation: Yulan Piao and Minoru Ko (National

Institute on Aging, NIH: <http://igsun.grc.nia.nih.gov/cDNA/>)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>Contact: nisc.mgc@nih.gov

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,

Blakesley, R.W., Bouffard, G., Breen, K., Brinkley, C., Brooks, S.,

Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,

Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lalic, P., Legaspi, R.,

Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, D.W.,

Tsurgenc, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,

Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAP Plate: 112 Row: 0 Column: 14

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis. Similarity but not identity to protein.

Location/Qualifiers

1..1493

/organism="Mus musculus"

/mol_type="mRNA"

/strains="C57BL/6NCR"

/db_xref="taxon:10090"

/clone="MGC:60769 IMAGE:30073287"

/tissue_type="Hematopoietic Stem Cell,"

/lin-/C-Kit-/Sca-1+), mouse, 10 wks"

/clone_lib="NIA Mouse Hematopoietic Stem Cell

/lin-/C-Kit-/Sca-1+ cDNA Library (Long)"

/lab_host="DH10B"

/note="Vector: pSPORT1"

1..1493

/gene="Ebi2"

/db_xref="LocusID:321019"

/db_xref="MGI:2442034"

73..1146

/codon_start=1

/product="Epstein-Barr virus induced gene 2"

/protein_id="AAH52868.1"

/db_xref="GI:31127265"

/db_xref="LocusID:321019"

/translation="MANNFTPLATSHNNCDLYAHHSARVLMPLHYSLVRLIGLV
NLLALVIVQKKNSTLYSNNLVISDLFTPLTRLYAYALGDFNRIGDALCRV
TALVFINIYAGVFNKISLDRFAVHPLRNKRIYARGVSLWVLYVFAQT
PLFUPMSKRGDKTTCHEYNFEGTASLPWILLGACILGYPLPITVILLVQYICCK
LFRATKONPLTEKSGVKNKALNTILIIIVFVILCFTFYHVALIIOHMKMLCFGALC
GARHSFQISLFTVCLNNFNCCMPFIYFFACKYKRVKMKMLKQVSVSISSAVRSA
PEESRENTESQMMIHSKASNGR"
202..984
misc_feature
/note="7cm_1; Region: 7 transmembrane receptor (rhodopsin
family)"
/db_xref="CDD:pfam00001"

misc_feature

ORIGIN
Query Match 19.9%; Score 220; DB 10; Length 1493;
Best Local Similarity 53.0%; Pred. No. 1.4e-34;
Matches 497; Conservative 0; Mismatches 435; Indels 6; Gaps 1;

ORIGIN

QY 101 CTCTGTTCTACACAGCCCTCTCTGGTCTTCACTGCGCCCTGGGAACATCTCTTGCC 160
DB 164 CTCTGATACAGCTGGTCTTCACTATGCGCTGGTGGGAACCTCTGCGCTTGGTGG 223
QY 161 TTACTGTCAAAGAGCAGGAGATCAACTGCAAGGCACTCTCTTCAATGAATGGTCA 283
DB 224 TCATTGTTCAAACAGAAAAAATCAACTCAACCACTCTTATTAATGAATGGTCA 283
QY 221 TGTCTGACTGCTGTTACCGTGGCTTACCGGAAGGCTGCTGTGTATGTCTGCTGGCT 280
DB 284 TTTCTGACATCTGTTTACACAGCTTACCACTCGATAGCTTACTATGCGCTGGCT 343
QY 281 CCACTGCGCTTTTCGGAAGGCTCTGACAGCTGACGGCTTGTCTCTACACGACA 340
DB 344 TTGATTTGAGGATAGTGTATCCCTGTGCGGGTAACTGCTGTGCTTCTTACATCA 403
QY 341 CCTAGCGGGGCTTACCTGATGCGCTGTGAGCGTGGACATTTACCAAGCTGTGCT 400
DB 404 CGTACGAGGTGTGAATCTATGCTTGTGAGCATAGACCGCTTCTTCTGCTGTGCT 463
QY 401 GTGCCCACTGGGCGCGCTTCCGACGGCTGCGCGCGCGCAGCTGGTGTGCTGCGCA 460
DB 464 ACCCTCTGCGCTTACCAAGATTAAGAAATCGAATACGCAAGGGTGTGCTGCTGCG 523
QY 461 TCTGACCTTGTGCTGTGACAGCATGCTTCTTGTATGCCCATGACCAAGCGCG 520
DB 524 TCTGATCTGCTTCTTGTCTCAACACTGCGCTGTTCTTCACTGCTTATGCTAAGGAG 583
QY 521 TGTGGGCAAGCTGCGCTGCTGATGAGTACAGCAGATGAGTCACTGCTGGGCTGCGCC 580
DB 584 AGGAGACAAAGACCACTTCTGATGGAGTATCCAAACTTTGAGGACAGCGTCCCTGCG 643
QY 581 TCATGCTCTGCTGCGCTTTCATGCTGCTTCTGCGGCGCAGTGGGATCATCTGCTCT 640
DB 644 GATTTCTGCTCGAGCTGCTGCTGCGGCTAGCTGCTGCTATCATCACTCTCTCTGT 703
QY 641 GCTATATGAGATCACTTGAAGCTGTGACAGCAGCTCGGAGGACCCAGTACACAGCA 700
DB 704 GTTACTCTCAGATCTGCTGCAAACTCTTTCAGGACTGCGCAAGCAGAACCTCAACG 763
QY 701 GGAAGGACGCCACTGCGGAGCTGCTGCTTATAGCTGCTGCTGCTGCTGCTGCTGCTG 760
DB 764 ATCTGCTGTGACAAAAGGCTCTCAACAATATCTCTTATCTCTGCTGCTGCTATCC 823
QY 761 TGTGCTTACGCCCTTACACACTCAACATCAAGCAGTTCATGCGAGGAGATGCTCACCC 820
DB 824 TGTGCTTACGCCCTTACACACTGCGGCTTCACTTACGACATGATAAAGATCTCTGCTCC 883
QY 821 TGCATATCC-----TGTGCGGAGCGGGGCTTCTTACTGCTGCTTTCAGGCCACCGTGG 874
DB 884 CTGAGCCCTTGGAGTGTGGGCGAGACATCTCTCCAGATCTCTCTGCACTTCAAGTGT 943
QY 875 CCTCATGACATGAGTGTGCGATACCCCAATCATTTACTTCTTTGCTGATCCACCATTT 934
DB 944 GCGTGTGAAGTCAACTGCTGCTGATGAGCCCGCTTATATATCTTCTTTCATGCAAGGGT 1003

TITLE

Direct Submission

QY 935 ACAGGAATGCTCTCTGGGCAATTTAAAGCTCAAAGGCTTCTCTCTCTCTCTCTCTCT 994
DB 1004 ATAAGAGAAGGTATGAAGATGCTCAAGCTCAAGTGTGTGTGTGTGTGTGTGTGT 1063
QY 995 CTCTCTCAACCCCAAGAAAAGCTTCTTCAGAAACACCA 1032
DB 1064 TGAGTCAAGCCCTGAAGAGAGTTCGCGGGAATGACA 1101

RESULT 12

AC133310/C

LOCUS

DEFINITION

AC133310

Rattus norvegicus

AC133310.2

HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

KEYWORDS

Rattus norvegicus

SOURCE

Rattus norvegicus

ORGANISM

Rattus

1 (bases 1 to 239344)

REFERENCE

AUTHORS

Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Amin, A., Anguiano, D., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Surrill, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, P., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, N., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabis, A., Ganta, R., Garcia, A., Garner, I., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guvera, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Liu, J., Kwis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, F., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuwa, L., Loulseg, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmood, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mayhew, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Pattertnak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Plopper, A., Popovic, D., Primus, E., Pu, L., Pu, L., Pu, L., Pu, L., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savory, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, K., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Weinstock, G., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission

QY 935 ACAGGAATGGCTCTGCGCATTTTAAAGCTCAAAAGGTTCTTCTCTCTCC 987
 DB 72209 ACAGGAAGGCTCATGAGATGCTCAAGGTCAGTGAGGCTGTCATCTCC 72157

RESULT 13
 AC094739/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-5C8, *** SEQUENCING IN PROGRESS ***.
 AC094739
 VERSION HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 244625)
 Muzny,D,Marie., Metzker,M.Lee., Abranzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Dwy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Haves,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Louisleged,H., Lozada,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangun,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,S., Nguyen,N., Norris,S., Parks,K., Nwackemeleh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Polindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rockey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,M., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission
 Unpublished
 2 (bases 1 to 244625)
 Worley,K.C.
 Direct Submission

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 244625)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On May 9, 2003 this sequence version replaced gi:2271707.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GBGO
 Center clone name: CH230-5C8
 ----- Summary Statistics
 Assembly program: Atlas;
 Consensus quality: 227319 bases at least Q40
 Consensus quality: 230441 bases at least Q30
 Consensus quality: 232587 bases at least Q20
 Estimated insert size: 238330; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 244625: contig of 244625 bp in length.

Location/Qualifiers
 1..244625
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-5C8"
 1..1336
 /note="wgs end extension
 clone_end:Sp6"
 1873..57771
 /note="clone boundary
 clone_end:Sp6
 site:ECORI
 end_sequence:BH335182"
 45659..47268
 /note="wgs contig"
 complement(242705..243403)
 /note="clone boundary
 clone_end:T7
 site:ECORI
 end_sequence:BH334721"

FEATURES
 source
 misc_feature
 misc_feature
 misc_feature
 misc_feature
 ORIGIN

Query Match 18.9%; Score 208.6; DB 2; Length 244625;
Best Local Similarity 53.0%; Pred. No. 3.4e-32;
Matches 473; Conservative 0; Mismatches 414; Indels 6; Gaps 1;

QY 101 CTCCTGTTCTACACAGCCCTCTCGTCTTCAGTGCCTCGGAAACATCTCTGCCCTTTGCC 160
DB 100918 CTCCTGATTATAGCTGGTCTTCATCTGGTCTCGTGGAAACATCTTTGGCCTTAGTTG 100859
QY 161 TTACCTGTCAAAAGACAGAGAGATCAACTGACAGGCACTCTACCTGGTGACCTGGGTG 220
DB 100858 TCATTGTTCAAAATAGAAAAAATCAACTCAACACTCTCTATTCAATGAACITGGTGA 100799
QY 221 TGTCTGACCTGCTGTCTACCGTGGCTTACCGGGAAGGTGTGTGTATGTCTGGCT 280
DB 100798 TTTCGCACATCTGTATTACACAGCTTTACCTACCGGATAGTCTACTATGCCCTGGGT 100739
QY 281 CCAGCTGGGCTTTCCGCAAGGGCTCTGACGCTCACGCGTGTGTGTCTACACCGACA 340
DB 100738 TTGACTGGAGGATTTGTTGATGCCCTGTGCGGATTAATCTGTCTGTCTTACATCA 100679
QY 341 CTTACGGGGGTCTACTCTCATGCTGTGTGAGCGTGGACCATTTACCAAGCTGTGGTCT 400
DB 100678 CGTATGCAGTGTGAATCTCATGCTGTGTGAGCATAGACGATCTTTGCGCTGGTGC 100619
QY 401 GTGCCACTGGGCGCGGCTCCGACGGCTGGCGCGCGAGGCTGTCTGGCTGGCCA 460
DB 100618 ACCCTCTGCGCTACAAACAGATTAAGAAATTAATATGAAGGCAATTTGTGTATTG 100559
QY 461 TCTGGACCTTGTGTCTGACAGATGCTCTCTCTGTATGCCATGACCAAGCGC 520
DB 100558 TCTGGATTTGTGTGTCTCAACACTGCACTGCTCTCAACCCCATGTCTAAGCAGG 100499
QY 521 TGGTGGCAAGCTGGCTGCTGAGTACAGCAGCATGAGTCACTCTCGGGTGGCCCC 580
DB 100498 AGCGACACAAGACTACTTGCATGGAATATCCAAATTTTGAAGGGAAGCATCTCTCCCGT 100439
QY 581 TCATGCTCTGTGGCTTTGGCATTTGCTGTGGCCAGTGGGGATCATCTCTGCTCT 640
DB 100438 GGATTTCTGCTCGAGCTGCTGCTAGCTAGCTGCTGCTCTTCCCATCATCTCTGCT 100379
QY 641 GCTATATGAAGATCACTGGAGCTGTGAGCAGCATGCTGGAGGACCCAGTGACAGCA 700
DB 100378 GCTACTCTCAGATCTGCTGCAAACTCTTCAGGACTGCCAAGCAGAACCCACTGACCGAGA 100319
QY 701 GGAAGAGCGCACTGGCGAGGCTGCTGTCTGCTGATGCTGTGTGGCGGTGGTG 760
DB 100318 AATCTGGTGTGAACAAAGGCTCTCAACAGTATCTCTCATCATCGCGGTGTCTGCT 100259
QY 761 TGTGCTTACGCCCCCTACCACTCAACATCAAGCATTTATGCGGAGAGGATGTCTCCACC 820
DB 100258 TGTGCTTACGCCCCCTACCACTGCGGCAATTTATGACGACATGTTGAAGACTCTCTACGCCC 100199
QY 821 TGCCATC-----CTGTGCGAGCGGAGGCTTTCTTACTGTCCTTTCAGGCCACCGTGG 874
DB 100198 CCGAGGCCCTCGGCTGTGGGGTGAGACATCTCTTCAGATCTCTCTGCACCTTCAAGGT 100139
QY 875 CCCTCATGAACATGAATCTGGCATTTACCCCAATCATTTACTTTCTTTCATCCACCCATT 934
DB 100138 GCTGATGAATCTCAACTCTGATGAGCCCTTTTCATATACTTTCTTTCATGTAAGGT 100079
QY 935 ACAGGAATGCTCTCGGCAATTTAAAGCTCAAGGCTCTCCCTCCCTCCCTCC 987
DB 100078 ACAAGAGAAGGTGATGAAGATGCTCAACAGCTCAAGTAGCGGTGTCAATCTCC 100026

RESULT 14

LOCUS AR308611
DEFINITION Sequence 77 from patent US 6555339.
ACCESSION AR308611
VERSION AR308611.1 GI:31700140
KEYWORDS
SOURCE Unknown.

ORGANISM Unknown.

Unclassified.

1 (bases 1 to 1086)

Liaw,C.W., Behan,D.P. and Chalmers,D.T.

Non-endogenous, constitutively activated human protein-coupled

receptors

Patent: US 655339-A 77 29-APR-2003;

Location/Qualifiers

source

1..1086

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Query Match 16.7%; Score 184.8; DB 6; Length 1086;
Best Local Similarity 50.6%; Pred. No. 2.1e-27;
Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;

QY 101 CTCGTCTTACACAGCCCTCTCGTCTTCAGTGCCTGGGAAACATCTCTGCCCTTTGCC 160
DB 104 CTCGTCAATTACAGCTCGTCTTCATCATTTGGGCTCGTGGAACTTACTAGCCTTGGTGC 163
QY 161 TTACCTGTCAAAAGACAGAGATCAACTGACAGGCACTTACCTGTGTGCACCTGGCTG 220
DB 164 TCATTGTTCAAAACAGAAAAAATCAACTTACCACCTCTATTCAACAAATTTGGTGA 223
QY 221 TGTCTGACCTGCTGTTCACCGTGGCTTACCGGGAAGGTGTGTGTATTATGTCTGGCT 280
DB 224 TTTCTGATATACTTTTACCAAGGCTTTTGCTTACAGATAGCTTACTATGCAATGGCT 283
QY 281 CCAGCTGGCCTTTCCGCAAGGGCTCTGACGGCTGACGGGCTTTGTGTCTTACACCGACA 340
DB 284 TTGACTGGAGATCGAGATGCCCTTGTAGGATACTGGCTAGTGTATTACATCAACA 343
QY 341 CTTACGGGGGTGTACTCTCATGCGCTGTGTGAGCGTGGACCATTTACCAGCTGTGGTCT 400
DB 344 CATATGCAAGTGTGAATCTTATGACCTGCTGAGTATTGACCGCTTCATTGTCTGGTGC 403
QY 401 GTGCCACTGGGCGCGGCTCCGACGCTGCGCGCGCAGGCTGGTCTGGTGGCCA 460
DB 404 ACCCTCTACCTACACAGATAAAAGATTTGAACATGCAAAAGCGGTGTGATATTG 463
QY 461 TGTGACCTTGTGTGTGTGACAGCATGCCCTTGTCTGTGATGCCCATGACCAAGCGC 520
DB 464 TCTGGATTTCTAGTATTGTCTGACACATCCCTCATCTCATCAACCTATGTCAAAGCAG 523
QY 521 TGTGGCAGCTGGCTGCTGATGAGTACAGCAGCATGAGTCACTCTCGGCTGCCCC 580
DB 524 AGCTGAAGATTAATGATGAGATTCCTCAACCTTTGAAGAACTAAATCTCTTCCCT 583
QY 581 TCATGCTCTGTGGCTTTGCCATTTGGCTTCTGTGGGCGAGTGGGGATCATCTGTCT 640
DB 584 GGATTTCTGCTGGGCTATTTTCTAGGATATGCTTCCACTTATAATCAATCTCTCATCT 643
QY 641 GCTATATGAAGATCACTGGAGCTGTGACGACAGCTCGGAGGACCCAGTGACCGACA 700
DB 644 GCTATTCTCAGATCTCTGCAAACTCTTCAAGACTGCCAAACAAACCCACTCCTGAGA 703
QY 701 GGAAGAGCGCACTGGCGAGGCTGCTGTTACGCTGTGATGCTGTGTGGCGCTGGTGG 760
DB 704 AATCTGGTGAACAAAGGCTCTCAACACAATATTCTTATTATTGTGTGTGTTTTC 763
QY 761 TGTGCTTACGCCCCCTACCACTCAACATCAAGCATTTATGCGGAGAGGATGCTCCACC 820
DB 764 TCTGTTTCAACCTTACCATTTGCAATTTTCAACATATGATTAAGAGCTTCGTTTCT 823
QY 821 TGCCATC-----TGTGCGAGCGGAGGCTTTCTTACTGTCCCTTCAGGCCACCGTGG 874
DB 824 CTAAATTTCTGGAATGTAGCAAGACATTCGTTCCAGATTTCTCTGCACCTTTACAGTAT 883
QY 875 CCCTCATGAACATGAATGTGGCATTTACCCCAATCATTTACTTTCTTGTGATCCACCAT 934
DB 884 GCCTGATGAATCTCAATTTGCTGATGAGCCCTTTTATCTACTTTCTTGTGATGTAAGGGT 943

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 17:32:38 ; Search time 484 Seconds

(without alignments)
9690.105 Million cell updates/sec

Title: US-10-041-615-33

Perfect score: 1104
Sequence: 1 atgataaacataggccctta.....gtgtcatttgcatttatgact 1104

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 28Jan04:*

- 1: Geneseqm1980s:*
- 2: Geneseqm1990s:*
- 3: Geneseqm2000s:*
- 4: Geneseqm2001as:*
- 5: Geneseqm2001bs:*
- 6: Geneseqm2002s:*
- 7: Geneseqm2003as:*
- 8: Geneseqm2003bs:*
- 9: Geneseqm2003cs:*
- 10: Geneseqm2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1104	100.0	1104	6	ABs64723 cDNA enco
2	779	70.6	4319	9	Adc86840 Human GPC
3	779	70.6	9409	4	Aak76477 Human Imm
4	672.8	60.9	1526	6	Abt94922 Human nov
5	655	59.3	1771	5	Aas78869 DNA enco
6	594.4	53.8	1001	6	Aas98131 Human DNA
7	397.2	36.0	536	4	Aas42840 Human G P
8	378.4	34.3	556	4	Aas30791 Human CDN
9	328.4	29.7	330	6	Aas98111 Human DNA
10	184.8	16.7	1086	3	Aaa30635 Human G p
11	184.8	16.7	1086	7	Acc48996 Human G-P
12	184.8	16.7	1086	9	Adc22596 Human G-P
13	184.8	16.7	1643	2	Aaq64126 Epsstein B
14	184.8	16.7	1643	3	Aav25491 cDNA for
15	184.8	16.7	1643	7	Acas56681 Human sig
16	184.8	16.7	1643	9	Ades4890 Farnesyl
17	184.8	16.7	2477	9	Ades25635 Human cDN
18	183.2	16.6	1086	3	Aaa30730 DNA enco
19	183.2	16.6	1086	9	Adc22724 Human G P
20	183.2	16.6	1638	7	Abz42660 Human Esv
21	172.2	15.6	1556	3	Aaa52629 Eosinophil
22	148	13.4	594	6	Aas98061 Human DNA
23	123.4	11.2	548	8	Ach23286 Human adu

ALIGNMENTS

RESULT 1

ABs64723
ID ABS64723 standard; cDNA; 1104 BP.

XX AC ABS64723;

XX AC ABS64723;

XX DT 15-NOV-2002 (first entry)

XX DE cDNA encoding human GPCR16 protein.

XX DE cDNA encoding human GPCR16 protein.

G protein coupled receptor related protein; ss; gene; human; GPCR;
cardiomyopathy; atherosclerosis; diabetes; cancer; stroke;
Von Hippel-Lindau syndrome; Alzheimer's disease; tuberculous sclerosis;
hypercalcaemia; Parkinson's disease; Huntington's disease;
cerebral palsy; epilepsy; Leach-Nyhan syndrome; multiple sclerosis;
ataxia-telangiectasia; leukodystrophy; addiction; anxiety; depression;
pain; obesity; Crohn's disease; osteoporosis; haemophilia; asthma;
inflammatory bowel disease; infertility; hypertension; scleroderma;
arthritis; human immunodeficiency virus; autoimmune disease; HIV;
infection; graft-versus-host disease.

XX OS Homo sapiens.

XX PN WO200264793-A2.

XX PD 22-AUG-2002.

XX PP 03-JAN-2002; 2002WO-US0000056.

XX PR 03-JAN-2001; 2001US-0259552P.

XX PR 09-JAN-2001; 2001US-0260544P.

XX PR 20-MAR-2001; 2001US-0277405P.

XX PA (CURA-) CURAGEN CORP.

XX PA Casman SJ, Edinger SR, Ellerman K, Smithson G, Kekuda R;

XX PI Murallidhara P;

XX XX WPI; 2002-643487/69.

XX DR P-PSDB; ABG32286.

XX PT New isolated G protein coupled receptor polypeptides and polynucleotides,
useful in gene therapy, particularly for treating or preventing
cardiomyopathy, atherosclerosis, diabetes, multiple sclerosis, Crohn's
disease or cancer.

24	119.4	10.8	1020	6	ABQ79300	Human GPC
25	119.4	10.8	1076	6	AAD29667	Human G-9
26	118.6	10.7	2025	4	AAD04980	Human pur
27	118.6	10.7	2025	6	ABZ35611	Human gen
28	118.6	10.7	2025	7	ACA56884	Human sig
29	118.6	10.7	2025	7	ABZ42713	Human pur
30	117	10.6	2118	6	ABZ35320	Human gen
31	116	10.5	1163	4	AAD04981	Human pur
32	115	10.4	1017	9	ADC25999	Human pur
33	115	10.4	1017	9	AAD61648	Human GPC
34	112	10.1	1020	3	AAA30601	Human G P
35	112	10.1	1020	3	AAA30718	DNA enco
36	112	10.1	1020	6	ABK52848	Human cys
37	112	10.1	1020	9	ADC22550	Human G P
38	112	10.1	1020	9	ADC22700	Human G P
39	112	10.1	1900	7	ACA56886	Human sig
40	112	10.1	1901	2	AAQ66178	Seven tra
41	112	10.1	1901	2	AAV18356	Human R12
42	112	10.1	1901	3	AAA91725	Human 7TM
43	112	10.1	1901	6	ABK54254	Human 7 t
44	112	10.1	2070	7	ACQ72705	Human can
45	112	10.1	2070	7	ABZ42757	Human G P

Claim 9; Page 97; 252pp; English.

This invention relates to a new isolated G-protein coupled receptor (GPCRX) polypeptide sequence and the cDNA encoding it. The GPCRX polypeptide, GPCRX nucleic acid and an antibody specific to the protein are useful for treating, preventing or alleviating a GPCRX-associated disorder or a pathological state in a subject, particularly a human. In particular, the disorder is cardiomyopathy, atherosclerosis, diabetes, or a disorder related to cell signal processing and metabolic pathway modulation. The GPCRX polypeptide and nucleic acid are also useful for diagnosing the presence of or predisposition to a disease associated with altered levels of GPCRX, particularly cancer. The GPCRX nucleic acid and polypeptide are especially useful in the manufacture of a medicament for therapeutic or prophylactic applications for disorders associated with aberrant GPCRX expression or activity, e.g. Von Hippel-Lindau syndrome, Alzheimer's disease, stroke, tuberosus sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, addiction, anxiety, depression, pain, obesity, Crohn's disease, osteoporosis, inflammatory bowel disease, infertility, hypertension, scleroderma, haemophilia, asthma, arthritis, human immunodeficiency virus, autoimmune disease, HIV, viral, fungal, bacterial or protozoal infections, or graft-versus-host disease. The DNA encoding the protein is useful in gene therapy for treating the above conditions. The polypeptides can be used as immunogens to produce antibodies and as vaccines. The nucleic acids are further used as hybridisation probes, in chromosome mapping, tissue typing, preventive medicine, and pharmacogenomics. These are also useful in developing powerful assay system for functional analysis of various human disorders, as well as in diagnostic applications. The present sequence represents the cDNA encoding a human G protein coupled receptor related protein (GPCR) of the invention

Sequence 1104 BP; 217 A; 328 C; 290 G; 269 T; 0 U; 0 Other;

Query Match 100.0%; Score 1104; DB 6; Length 1104;
Best Local Similarity 100.0%; Pred. No. 6.8e-254;
Matches 1104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATAAAGTAGGCGCTTACTTACACACCCACCAAAATAAACTAAGATTAAAGAT 60
DB 1 ATGATAAAGTAGGCGCTTACTTACACACCCACCAAAATAAACTAAGATTAAAGAT 60
QY 61 TTAATATGGAAGCAAACTGTGAAATTTTGGCTCTCTGTCTACACAGCCCTC 120
DB 61 TTAATATGGAAGCAAACTGTGAAATTTTGGCTCTCTGTCTACACAGCCCTC 120
QY 121 CTGGCTTTCAGTCCCTGGGAAACATCTCTGCTTTCCTTACCTGTCAAAAGAGCAG 180
DB 121 CTGGCTTTCAGTCCCTGGGAAACATCTCTGCTTTCCTTACCTGTCAAAAGAGCAG 180
QY 181 AAGATCAACTGCACAGGATCTACCTGTGCACCTGTGTCTGTACCTGTCTTCAAC 240
DB 181 AAGATCAACTGCACAGGATCTACCTGTGCACCTGTGTCTGTACCTGTCTTCAAC 240
QY 241 GTGGCTTTCAGGGAAGGCTGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
DB 241 GTGGCTTTCAGGGAAGGCTGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
QY 301 GGGCTCTGAGCTGACCGCTTGTGTCTTACACCAACCTACACGGGGGGTCTACCTC 360
DB 301 GGGCTCTGAGCTGACCGCTTGTGTCTTACACCAACCTACACGGGGGGTCTACCTC 360
QY 361 ATGGCTGTGTGAGCTGACCATTTACCCAGCTGTGGTCTGTGCTGTGCTGTGCTGTGCT 420
DB 361 ATGGCTGTGTGAGCTGACCATTTACCCAGCTGTGGTCTGTGCTGTGCTGTGCTGTGCT 420
QY 421 CTCGCAAGCTGTGCGGCGGCAAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
DB 421 CTCGCAAGCTGTGCGGCGGCAAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
QY 481 CAGACGATCCCTTCTCTTGTATGCCATGACCAAGCGCTGTGGGCAAGCTGGGCTGC 540

DB 481 CAGACGATCCCTTCTCTTGTATGCCATGACCAAGCGCTGTGGGCAAGCTGGGCTGC 540
QY 541 ATGAGATACAGAGCATGAGTCACTCTGGGCTGCCCTCATGGTCTCTGTGGCTTTT 600
DB 541 ATGAGATACAGAGCATGAGTCACTCTGGGCTGCCCTCATGGTCTCTGTGGCTTTT 600
QY 601 GCCATTGGCTTCTGTGGGCGAGTGGGATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 660
DB 601 GCCATTGGCTTCTGTGGGCGAGTGGGATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 660
QY 661 AAGCTGTGAGCAAGCTGTGGGAGGACCCAGTACACAGAGGAAAGGACGACCTGGCGA 720
DB 661 AAGCTGTGAGCAAGCTGTGGGAGGACCCAGTACACAGAGGAAAGGACGACCTGGCGA 720
QY 721 GGCTGCTGCTTACGCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780
DB 721 GGCTGCTGCTTACGCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780
QY 781 CTCACATCAAGCAGTTTATGGGAGGAGTCTCCACTGCCATCTCTGTGCGGAGCGG 840
DB 781 CTCACATCAAGCAGTTTATGGGAGGAGTCTCCACTGCCATCTCTGTGCGGAGCGG 840
QY 841 AGGCTTCTTACTGTCCCTTCAAGGCGACCGTGGCCCTCATGAACATGAATCTGTGCAAT 900
DB 841 AGGCTTCTTACTGTCCCTTCAAGGCGACCGTGGCCCTCATGAACATGAATCTGTGCAAT 900
QY 901 ACCCAATCAATTTACTTTTGTATCCATCCATCAGGAAATGGCTCTGTGGCATTTTA 960
DB 901 ACCCAATCAATTTACTTTTGTATCCATCCATCAGGAAATGGCTCTGTGGCATTTTA 960
QY 961 AAGCTCAAGGGTCTTCT 1020
DB 961 AAGCTCAAGGGTCTTCT 1020
QY 1021 TCAGAAACACCAAGTATACCCAGGAGGATCTCCACTGCCATCTCTGTGCGGAGCGG 1080
DB 1021 TCAGAAACACCAAGTATACCCAGGAGGATCTCCACTGCCATCTCTGTGCGGAGCGG 1080
QY 1081 TAACTGTCTATTGCTTTATGACT 1104
DB 1081 TAACTGTCTATTGCTTTATGACT 1104
RESULT 2
ADC86840
ID ADC86840 standard; DNA; 4319 BP.
XX AC ADC86840;
XX AC ADC86840;
DT 01-JAN-2004 (first entry)
XX Human GPCR gene SEQ ID NO:1293.
DE ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX Homo sapiens.
XX EPI270724-A2.
XX 02-JAN-2003.
XX 18-JUN-2002; 2002EP-00013517.
XX 18-JUN-2001; 2001JP-00246789.
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX Suwa M, Asai K, Akiyama Y, Aburatani H;
XX WPI; 2003-315783/31.
XX P-PSDB; ADC86841.
DR

XX New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 1293; 289p; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC8548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 4319 BP; 1058 A; 1049 C; 1170 G; 994 T; 0 U; 48 Other;
Query Match 70.6%; Score 779; DB 9; Length 4319;
Best Local Similarity 83.6%; Pred. No. 5.6e-176; Mismatches 0; Indels 197; Gaps 3;
Matches 1006; Conservative 0;
97 CGCTCTCTGTTCTACAGCCCTCTGCTCTTCAAGTCCCTGGGAAACATCTTGCCCTT 156
Db CGCTCTCTGTTCTACAGCCCTCTGCTCTTCAAGTCCCTGGGAAACATCTTGCCCTT 2999
97 TGCCTTACTGTCAAAAGACGAGGAGATCAACTGCACAGGATCTACTGTCACCTG 216
Db TGCCTTACTGTCAAAAGACGAGGAGATCAACTGCACAGGATCTACTGTCACCTG 3059
217 GCTGTGTCTGACTCTGTTTCACTGCTGCTTACCGGAAAGGCTGCTGTTATGTGCTG 276
Db GCTGTGTCTGACTCTGTTTCACTGCTGCTTACCGGAAAGGCTGCTGTTATGTGCTG 3119
277 GCTCAGCTGCTGCTTTCGCAAGGGGCTCTGAGGCTGACGGGCTTGTGCTCTACACC 336
Db GCTCAGCTGCTGCTTTCGCAAGGGGCTCTGAGGCTGACGGGCTTGTGCTCTACACC 3179
337 GACACTGACGGGGGCTCTACCTCATGCTGCTGAGCGTGGACATTTACCAGCTGTG 396
Db GACACTGAC-GGGGGGCTCTACCTCATGCTGCTGAGCGTGGACATTTACCAGCTGTG 3238
397 GTCTGTGCCACTGGGGCCCGCTTCGCAAGGCTGCGCGCGCAGGCTGCTGTGCGTG 456
Db GTCTGTGCCACTGGGGCCCGCTTCGCAAGGCTGCGCGCGCAGGCTGCTGTGCGTG 3298
457 GCATCTGACCTTGGTGTGCTGTCAGACGATGCCCTTGTCTTGATGCCCATGACCAAG 516
Db GCATCTGACCTTGGTGTGCTGTCAGACGATGCCCTTGTCTTGATGCCCATGACCAAG 3358
517 CGCTGGTGGGCAAGCTGGCTGCATGGAGTACAGCAGCATGGAGTCACTCTCGGGCTG 576
Db CGCTGGTGGGCAAGCTGGCTGCATGGAGTACAGCAGCATGGAGTCACTCTCGGGCTG 3418
577 CCCCTCATGGTCTGCTGGTGGCTTTGCCATGSCCTTCTGTGGCCAGTGGGATCATCTG 636
Db CCCCTCATGGTCTGCTGGTGGCTTTGCCATGSCCTTCTGTGGCCAGTGGGATCATCTG 3478
637 TCTGTCTATCAAGATCACTTGGAGCTGTGCAGCAGCTT----- 678
Db TCTGTCTATCAAGATCACTTGGAGCTGTGCAGCAGCTTGGGAGAACCCAGTGACCA 3538
679 ----- 678
3539 GCGGGAAAGGACACACCGCGCGGCGAGCCAGGAGACCCAGTGACCAACAGGAAAGGA 3598
679 ----- 678
3599 CGCCACCGCGGGGAGGCCAGGAGGACCCAGTGACCGAGCGGAAAGACACCCCGGCA 3658
679 -----CGG 681
3659 GGACAGCCCGAGAGGACCCAGTGACCGAGCGGAAAGGACCGCAGCGGGGAGACCCCGG 3718

QY 682 GAGGACCCAGTACACGACGAGGAAAGGACGCCACTGCGAGGCTGCTGCTTACGTCGTG 741
Db GAGGACCCAGTACACGACGAGGAAAGGACGCCACTGCGAGGCTGCTGCTTACGTCGTG 3778
QY 742 ATGCTGTGTCGCTGCTGCTGCTTACGCCCTTACGCCCTCAACATCAAGCAGTTTCATG 801
Db ATGCTGTGTCGCTGCTGCTGCTTACGCCCTTACGCCCTCAACATCAAGCAGTTTCATG 3838
QY 802 GCGAGAGGATGTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 861
Db GCGAGAGGATGTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3898
QY 862 CAGGCCACGCTGGCCCTCATGACATGACATGACATGACATGACATGACATGACATGACAT 921
Db CAGGCCACGCTGGCCCTCATGACATGACATGACATGACATGACATGACATGACATGACAT 3957
QY 922 GCATCCACCCATTACAGGAAATGGCTCTCTGGGCAATTTAAAGCTCAAGGGTCTTCCTCC 981
Db GCATCCACCCATTACAGGAAATGGCTCTCTGGGCAATTTAAAGCTCAAGGGTCTTCCTCC 4017
QY 982 TCT 1041
Db TCT 4077
QY 1042 CAGGCCAGAGGTTCTATGTTTCTTAGCGGACGCTGCTTAAAGCTCAAGGGTCTTCCTCC 1101
Db CAGGCCAGAGGTTCTATGTTTCTTAGCGGACGCTGCTTAAAGCTCAAGGGTCTTCCTCC 4137
QY 1102 ACT 1104
Db 4138 ACT 4140

RESULT 3

AAK76477
ID AAK76477 standard; DNA; 9409 BP.
XX AAK76477;
XX
XX 07-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:31289.
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
KW Homo sapiens.
XX
XX WO200157182-A2.
PN 09-AUG-2001.
PD
PP 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180528P.
PR 24-FEB-2000; 2000US-0184564P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 18-APR-2000; 2000US-0190076P.
PR 19-MAY-2000; 2000US-0205151P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.

PR	08-NOV-2000;	2000US-02464778.	
PR	08-NOV-2000;	2000US-0246478P.	
PR	08-NOV-2000;	2000US-0246523P.	
PR	08-NOV-2000;	2000US-0246524P.	
PR	08-NOV-2000;	2000US-0246525P.	
PR	08-NOV-2000;	2000US-0246526P.	
PR	08-NOV-2000;	2000US-0246527P.	
PR	08-NOV-2000;	2000US-0246528P.	
PR	08-NOV-2000;	2000US-0246532P.	
PR	08-NOV-2000;	2000US-0246560P.	
PR	08-NOV-2000;	2000US-0246610P.	
PR	08-NOV-2000;	2000US-0246611P.	
PR	08-NOV-2000;	2000US-0246613P.	
PR	17-NOV-2000;	2000US-0249207P.	
PR	17-NOV-2000;	2000US-0249208P.	
PR	17-NOV-2000;	2000US-0249209P.	
PR	17-NOV-2000;	2000US-0249210P.	
PR	17-NOV-2000;	2000US-0249211P.	
PR	17-NOV-2000;	2000US-0249212P.	
PR	17-NOV-2000;	2000US-0249213P.	
PR	17-NOV-2000;	2000US-0249214P.	
PR	17-NOV-2000;	2000US-0249215P.	
PR	17-NOV-2000;	2000US-0249216P.	
PR	17-NOV-2000;	2000US-0249217P.	
PR	17-NOV-2000;	2000US-0249218P.	
PR	17-NOV-2000;	2000US-0249244P.	
PR	17-NOV-2000;	2000US-0249245P.	
PR	17-NOV-2000;	2000US-0249264P.	
PR	17-NOV-2000;	2000US-0249265P.	
PR	17-NOV-2000;	2000US-0249297P.	
PR	17-NOV-2000;	2000US-0249298P.	
PR	17-NOV-2000;	2000US-0249299P.	
PR	01-DEC-2000;	2000US-0249300P.	
PR	01-DEC-2000;	2000US-0249316P.	
PR	01-DEC-2000;	2000US-0250391P.	
PR	05-DEC-2000;	2000US-0250103P.	
PR	05-DEC-2000;	2000US-0251198P.	
PR	05-DEC-2000;	2000US-0256719P.	
PR	06-DEC-2000;	2000US-0256719P.	
PR	08-DEC-2000;	2000US-0251856P.	
PR	08-DEC-2000;	2000US-0251868P.	
PR	08-DEC-2000;	2000US-0251869P.	
PR	08-DEC-2000;	2000US-0251989P.	
PR	08-DEC-2000;	2000US-0251990P.	
PR	11-DEC-2000;	2000US-0254097P.	
PR	05-JAN-2001;	2001US-0259578P.	
XX			
XX	(HUMA-)	HUMAN GENOME SCI INC.	
PA			
XX			
PI	Rosen CA, Barash SC, Ruben		
XX	WPI; 2001-483426/52.		
DR			

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis.

Disclosure; SEQ ID NO 31289; 307ipp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/haematopoietic-related diseases, especially cancers and cancer metastases of haematopoietic-derived cells. AAK64703 to AAK67694 represent human immune/haematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAM82169

TK _____

CC sclerosis. The sequences are involved in chemotactic or chemokinetic
CC activity, regulation of hematopoiesis, treatment of myeloid or lymphoid
CC cell disorders and platelet disorders such as thrombocytopenia,
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC disease. The sequences of the invention are also useful for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, immune deficiencies and disorders
CC including severe combined immunodeficiency (SCID), bacterial or fungal
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC and coagulation disorders. Sequences ABK94590-ABK94592 represent human
CC novel polynucleotides of the invention
XX
SQ Sequence 1526 BP; 282 A; 458 C; 458 G; 328 T; 0 U; 0 Other;

Query Match 60.9%; Score 672.8; DB 6; Length 1526;
Best Local Similarity 81.9%; Pred. No. 9.7e-151;
Matches 901; Conservative 0; Mismatches 2; Indels 197; Gaps 3;

QY 98 GCTCTCTGTTTACACAGCCCTCTGCTTTCAGTGGCCCTGGGAACATCTTGGCCCTTT 157
DB |||||||
QY 158 GCTTACCTGTCAAAGAGCAGGAGATCAACTGCACAGGATCTACTGTGTACCTGG 217
DB |||||||
QY 482 GCTTACCTGTCAAAGAGCAGGAGATCAACTGCACAGGATCTACTGTGTACCTGG 541
DB |||||||
QY 218 CTGTCTCTGCTGTCTACCTGCTTACCTGCTTACCTGCTTACCTGCTTACCTGCT 277
DB |||||||
QY 542 CTGTCTCTGCTGTCTTACCTGCTTACCTGCTTACCTGCTTACCTGCTTACCTGCT 601
DB |||||||
QY 278 GCTCAGCTGCTCTTTCGCGAAGGGCTCTGCGAGCTGACGGCTTGTGCTTACACCG 337
DB |||||||
QY 602 GCTCAGCTGCTCTTTCGCGAAGGGCTCTGCGAGCTGACGGCTTGTGCTTACACCG 661
DB |||||||
QY 338 ACACCTAGCGGGGCTTACCTGCTTACCTGCTTACCTGCTTACCTGCTTACCTGCT 397
DB |||||||
QY 662 ACACCTAC- GGGGGGCTTACCTGCTTACCTGCTTACCTGCTTACCTGCTTACCTGCT 720
DB |||||||
QY 398 TCTGTGCCACTGGGGCCCGCTTCCGCAAGGGCTCTGCGAGCTGACGGCTTGTGCTTAC 457
DB |||||||
QY 721 TCTGTGCCACTGGGGCCCGTCTTCCGCAAGGGCTCTGCGAGCTGACGGCTTGTGCTTAC 780
DB |||||||
QY 458 CCATCTGACCTTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 517
DB |||||||
QY 781 CCATCTGACCTTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB |||||||
QY 518 GCTGTGGGCAAGCTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 577
DB |||||||
QY 841 GCTGTGGGCAAGCTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
DB |||||||
QY 578 CCCTCATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 637
DB |||||||
QY 901 CCCTCATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
DB |||||||
QY 638 CTTGCTATATGAATCATCTGGAAGCTGTGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCT 678
DB |||||||
QY 961 CTTGCTATATGAATCATCTGGAAGCTGTGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
DB |||||||
QY 679 ----- 678
DB 1021 CGGGAAGGACACACACCGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
QY ----- 678
DB 1081 GCGACCGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
QY ----- CGGG 682
DB 1141 GACAGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200

QY 683 AGGACCCAGTGCACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 742
DB |||||||
QY 1201 AGGACCCAGTGCACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
DB |||||||
QY 743 TGCTGTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 802
DB |||||||
QY 1261 TGCTGTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
DB |||||||
QY 803 CGAGAGGAGTGTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 862
DB |||||||
QY 1321 CGAGAGGAGTGTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB |||||||
QY 863 AGGACCCAGTGTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 922
DB |||||||
QY 1381 AGGACCCAGTGTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1439
DB |||||||
QY 923 CATCCACCCATTACAGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 982
DB |||||||
QY 1440 CATCCACCCATTACAGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1499
DB |||||||
QY 983 CTTCT 1002
DB |||||||
QY 1500 CTTCT 1519
DB |||||||

RESULT 5

AAS78869

ID AAS78869 standard; cDNA; 1771 BP.

XX AAS78869;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #14673.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG14682.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 1; SEQ ID NO 14673; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging

CC antibodies raised against the polypeptides, and methods of identifying
CC modulators of the polypeptides. the polypeptides are useful for
CC identifying modulator compounds which function as modulators, activators,
CC repressors, agonists or antagonists of the novel GPCR polypeptides
CC including the GAL4 polypeptide. The antibodies and nucleic acid probes as
CC described above can be used to detect the presence of the polypeptides
CC and nucleic acids and are used to diagnose a variety of diseases or
CC disorders in which GPCRs are involved e.g., Alzheimer's disease,
CC amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell
CC carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, chronic
CC obstructive pulmonary disease, Crohn's disease, depression, epilepsy,
CC macular degeneration, lymphoma, melanoma, multiple sclerosis,
CC osteoarthritis, osteoporosis, Parkinson's disease, psoriasis, rheumatoid
CC arthritis, schizophrenia, ulcerative colitis, tuberculosis and many other
CC diseases listed in the specification. The probes and antibodies are also
CC useful for diagnosing cognition and memory disorders, anorexia, hormonal
CC release disorders, cardiovascular activity disorders, pain perception
CC disorders, obesity, diabetes, Alzheimer's disease. Preferably, compounds
CC that decrease or increase the expression of galanin receptor (GAL4) can
CC be used to treat obesity, diabetes, hyperlipidemia and stroke. The GPCR
CC nucleic acid is useful for treating the above mentioned disorders by gene
CC therapy techniques. The present sequence is a novel GPCR polynucleotide
CC of the invention
XX
SQ Sequence 1001 BP; 190 A; 315 C; 311 G; 195 T; 0 U; 0 Other;

Query Match		53.8%;	Score 594.4;	DB 6;	Length 1001;
Best Local Similarity		98.7%;	Pred. No. 4.7e-132;		
Matches 620;		Conservative	0;	Mismatches	6;
				Indels	2;
				Gaps	2;
QY	100	TCCTCTTTTACACAGCCCTCTGCTCTCACTGCGCCCTGGGAAACATCTTGCCCTTTGC	159		
DB	1001	TCCTCTTTTACACAGCCCTCTGCTCTCACTGCGCCCTGGGAAACATCTTGCCCTTTGC	942		
QY	160	CTTACCTGTCAAAGACAGGAGAACTCACTGCACAGGCATCTACTGCTGCACCTGCT	219		
DB	941	CTTACCTGTCAAAGACAGGAGAACTCACTGCACAGGCATCTACTGCTGCACCTGCT	882		
QY	220	GTGCTGTGACCTGTCTTCACTGCTGCTTACCGGAGGCTGTGTGTGTGTGTGTGTGTGT	279		
DB	881	GTGCTGTGACCTGTCTTCACTGCTGCTTACCGGAGGCTGTGTGTGTGTGTGTGTGTGT	822		
QY	280	TCAGCTGGCTTTTCGGCAAGGGCTCTGCAGGCTGACGGCTGTTGTGTGTGTGTGTGTGT	339		
DB	821	TCAGCTGGCTTTTCGGCAAGGGCTCTGCAGGCTGACGGCTGTTGTGTGTGTGTGTGTGT	762		
QY	340	ACCTACGGGGGGTCTACCTCTATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	399		
DB	761	ACCTAC -GGGGGGTCTACCTCTATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	703		
QY	400	TGTGCCCACTGGGGCCGGCCCTCCGACGGCTGGCCGGCCGGCTGTGTGTGTGTGTGTGT	459		
DB	702	TGTGCCCACTGGGGCCGGCCCTCCGACGGCTGGCCGGCCGGCTGTGTGTGTGTGTGTGT	643		
QY	460	ATCTGGACCTTGT	519		
DB	642	ATCTGGACCTTGT	583		
QY	520	CTGTGTGGCAAGCTGGGCTGCTGAGTACAGCAGCATGAGTCACTCTGGGCTGGCC	579		
DB	582	CTGTGTGGCAAGCTGGGCTGCTGAGTACAGCAGCATGAGTCACTCTGGGCTGGCC	523		
QY	580	CTCATGCTCTGT	639		
DB	522	CTCATGCTCTGT	463		
QY	640	TGCTATATGAGATCACTGTGAAGCTGTGACGACAGCTCGGAGGAGCCAGTACACG	699		
DB	462	TGCTATATGAGATCACTGTGAAGCTGTGACGACAGCTCGGAGGAGCCAGTACACG	404		
QY	700	AGGAAGGAGCCACTGCGAGGCTGCC	727		
DB	403	GGGAAGGAGCACACCGCGGGGAGCC	376		

RESULT 7
AAS42840
ID AAS42840 standard; cDNA; 536 BP.
XX
AC AAS42840;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human G Protein-Coupled Receptor (GPCR) cDNA #35.
XX
KW Human; G-protein coupled receptor; GPCR; mental disorder; schizophrenia;
KW attention deficit disorder; anxiety; depression; bipolar disorder; ss;
KW neurological disorder; Huntington's disease; dementia; obesity; anorexia;
KW metabolic disorder; Parkinson's disease; Tourette's syndrome; thrombosis;
KW type 2 diabetes; cardiovascular disorder; myocardial infarction; cancer;
KW cardiomyopathy; atherosclerosis; human immunodeficiency virus; HIV;
KW viral infection; immunostimulant; neuroleptic; nootropic; tranquiliser;
KW antidepressant; anorectic; PCR primer; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200162797-A2.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-US005676.
XX
PR 23-FEB-2000; 2000US-0184247P.
PR 23-FEB-2000; 2000US-0184303P.
PR 23-FEB-2000; 2000US-0184304P.
PR 23-FEB-2000; 2000US-0184305P.
PR 23-FEB-2000; 2000US-0184397P.
PR 02-MAR-2000; 2000US-0186457P.
PR 03-MAR-2000; 2000US-0186810P.
PR 09-MAR-2000; 2000US-0188064P.
PR 13-MAR-2000; 2000US-0188880P.
PR 03-APR-2000; 2000US-0194344P.
PR 23-JUN-2000; 2000US-0213861P.
PR 11-JUL-2000; 2000US-0217369P.
PR 11-JUL-2000; 2000US-0217370P.
PR 14-JUL-2000; 2000US-0218337P.
PR 20-JUL-2000; 2000US-0218492P.
XX
PA (PHAA) PHARMACIA & UPJOHN CO.
XX
PI Vogeli G, Wood LS, Parodi LA, Lind P;
XX
XX WPI; 2001-570628/64.
XX P-PSDB; AAU25588.
XX
XX New isolated nucleic acid encoding a new G-protein coupled receptor
XX polypeptide for detecting receptor modulators that can treat mental
XX disorders, such as schizophrenia, anxiety, depression, or obesity.
XX
XX Claim 4; Page 82-83; 279pp; English.
XX
XX Sequences AAS42806-AAS42926 represent cDNA molecules and PCR primers for
XX cDNA molecules encoding human G-protein coupled receptor (GPCR)
XX polypeptides. The protein and DNA sequences of the invention can be used
XX to identify compounds which bind to GPCR polypeptides and in screening
XX for compounds that modulate GPCR activity. By screening a human subject
XX for the presence of mutations in GPCR DNA, a GPCR-related disorder or a
XX genetic predisposition can be diagnosed. The sequences can also be used
XX for treatment and prevention of mental disorders such as schizophrenia,
XX attention deficit disorder, anxiety, depression, dementia and bipolar
XX disorder, neurological disorders such as Huntington's disease,
XX Parkinson's disease and Tourette's syndrome, metabolic disorders such as
XX obesity, anorexia and type 2 diabetes, cardiovascular disorders such as
XX thrombosis, myocardial infarction, cancers
XX viral infections caused by HIV and cancers

Seq	Sequence	536 BP; 93 A; 166 C; 182 G; 95 T; 0 U; 0 Other;
Query Match	36.0%;	Score 397.2; DB 4; Length 536;
Best Local Similarity	97.5%;	Pred. No. 5.3e-85;
Matches 435;	Conservative 0;	Mismatches 8; Indels 3; Gaps 3;
QY	285	CTGGCCCTTTGGCAAGGGGCTCTGAGGCTGACGGCGTTTGTGCTCTACACGACACCTA 344
DB	8	CTGGCCCTTTGGCAAGGGGCTCTGAGGCTGACGGCGTTTGTGCTCTACACGACACCTA 67
QY	345	CGGGGGGGTCTACCTCATGGCTGTGTGAGCGTGGACCAATTACCCAGCTGTGTCTGTGC 404
DB	68	C-GGGGGGTTCTACCTCATGGCTGTGTGAGCGTGGACCAATTACCCAGCTGTGTCTGTGC 126
QY	405	CCACTGGGGCCCGCGCCCTCCGACAGGCTGGCCGCGCCAGAGCTGGTCTGTGTGGCCATCTG 464
DB	127	CCACTGGGGCCCGCGCCCTCCGACAGGCTGGCCGCGCCAGAGCTGGTCTGTGTGGCCATCTG 186
QY	465	GACCTTGGTGTGCTGTCGACAGATGGCCCTTCTCTTCATGCCATGACCAAGCCGCTGGT 524
DB	187	GACCTTGGTGTGCTGTCGACAGATGGCCCTTCTCTTCATGCCATGACCAAGCCGCTGGT 246
QY	525	GGCGAAGCTGGCCTGCATGGAGTACAGCAGCATCGAGTCACTCTC-GGGCTGCCCTCA 583
DB	247	GGCGAAGCTGGCCTGCATGGAGTACAGCAGCATCGAGTCACTCTC-GGGGCTGCCCTCA 306
QY	584	TGGTCTGTGTGGCTTTGGCATGGCTCTGTGGCCAGTGGGGATCATTCTGTCTCTGCT 643
DB	307	TGGTCTGTGTGGCTTTGGCATGGCTCTGTGGCCAGTGGGGATCATTCTGTCTCTGCT 366
QY	644	ATATGAAGATCACTGGGAAGCTGTGCAGCAGAGCTCGGGAGGACCCAGTGACACGAGA 703
DB	367	ATATGAAGATCACTGGGAAGCTGTGCAGCAGAGCTCGGGAGGACCCAGTGACACGAGA 425
QY	704	AGGACGCCACTGGCGAGGCTGCCTG 729
DB	426	AAGGACACCAACCGCGGGCGAGCCCG 451
RESULT 8		
AA830791/c		
ID	AA830791	standard; cDNA; 556 BP.
XX	AA830791;	
AC		
XX		
DT	04-DEC-2001	(first entry)
XX		
DE	Human cDNA encoding G protein-coupled receptor nGPR-2343.	
XX		
KW	Human; G protein-coupled receptor; nGPR-x; ss; antiviral; analgesic;	
KW	cytostatic; cardiant; antidiabetic; anorectic; hypotensive; hypertensive;	
KW	antiparkinsonian; nootropic; neuroprotective; antidepressant;	
KW	viral infection; HIV-1; human immunodeficiency virus; HIV-2; pain;	
KW	cancer; metabolic disease; cardiovascular disease; type 2 diabetes;	
KW	obesity; anorexia; hypotension; hypertension; myocardial infarction;	
KW	atherosclerosis; parkinson's disease; psychosis; neurological disorder;	
KW	schizophrenia; migraine; major depression; anxiety; mental disorder;	
KW	manic depression; dykinesia; Huntington's disease; Tourette's Syndrome.	
OS	Homo sapiens.	
XX		
FN	W0200166750-A2.	
XX		
PD	13-SEP-2001.	
XX		
PF	08-MAR-2001; 2001WO-US0007322.	
XX		
PR	08-MAR-2000; 2000US-0187581P.	
PR	08-MAR-2000; 2000US-0187582P.	
PR	08-MAR-2000; 2000US-0187714P.	
PR	08-MAR-2000; 2000US-0187715P.	
PR	08-MAR-2000; 2000US-0187825P.	
PR	08-MAR-2000; 2000US-0187828P.	

of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
CC and AAA30775-A30779). The mutant proteins of the invention contain a
CC mutation in a portion of the protein comprising intracellular loop 3
CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC is substituted for an endogenous residue in IC3 at a position 16 amino
CC acids N-terminal of an endogenous proline in TM6 to form a sequence X-
CC (AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg or
CC Ala, and is preferably Lys. When the endogenous residue at this position
CC is Lys, this residue is replaced by His, Arg or preferably Ala. The 15
CC amino acid stretch between the substituted amino acid and the Pro may be
CC endogenous, non-endogenous, or a mixture of endogenous and non-endogenous
CC residues. The constitutively active GPCRs are useful for identifying
CC antagonists, agonists and partial agonists for use as pharmaceutical
CC agents. The mutant proteins are also useful in research settings for
CC elucidating the roles of the receptors in normal and diseased conditions.
CC Antagonists for a particular GPCR are useful for treating diseases and
CC disorders associated with that receptor. Because the novel mutant GPCRs
CC are constitutively active, they can be used directly for screening of
CC compounds without the need for endogenous ligands. The present sequence
CC represents cDNA encoding a human wild-type GPCR used in an
CC exemplification of the invention. This was cloned and subjected to site-
CC directed mutagenesis (SDM) to generate DNA encoding the corresponding
CC mutant of the invention
XX
SQ Sequence 1086 BP; 314 A; 242 C; 206 G; 324 T; 0 U; 0 Other;

Query Match 16.7%; Score 184.8; DB 3; Length 1086;
Best Local Similarity 50.6%; Pred. No. 4.1e-34;
Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;
101 CTCTGTTCTACACAGCCCTCTGCTGCTTTCAGTGCCTCGGAAACATCTTCGCCCTTTGCC 160
DB 104 CTCTGCATTACAGCCCTGCTTTCATCATTTGGCTCTGTGGAACTTACTAGCCTTGGTCG 163
QY 161 TTACCTGTCTAAAGAGCAGAGATCACTGACAGGATCTTACCTGTGTGACCTGGCTG 220
DB 164 TCATTGTTCAAAACAGGAAATAAACACTTACCACCTCTATTCAACAAATTGGTGA 223
QY 221 TGTCTGACCTGCTTTCACCTGGCTTACCCTGAGGAGGTGTGTATTGCTGTGGCT 280
DB 224 TTCTGATATACCTTTTACACGGCTTTCCTACAGGATAGCTTACTATGCAATGGCT 283
QY 281 CCAGCTGGCTTTTGGCAAGGGGCTCTGACGCTGACGGCTTTGTGCTCTACACCGACA 340
DB 284 TTGACTGGAGATCGGAGATGCTTGTGTAGGATAACTCGCTAGTGTTTTACATCAACA 343
QY 341 CCTACGGGGGGTCTACCTCATGGCTGTGTGAGGTGACCATTACCAGCTGTGGTCT 400
DB 344 CATATGAGTGTGAACCTTATGACCTGCTGAGTATGACCGCTTCTGCTGTGGTGC 403
QY 401 GTCCCACTGGGGCCCGCCTCCGACCGCTGGCGCGCCAGGCTGTCTCGTGGCCA 460
DB 404 ACCCTCTACGTACAACAAGATAAAAAGGATGAACATGCAAAAGCGGTGTGCATATTG 463
QY 461 TCTGGACCTTGTGCTGCTGCAGCATGCTTCTGCTTGTATGCCATGACCAAGCCGC 520
DB 464 TCTGGATTTAGTATTGTGCTACAGACTCCCACTCTCTCATCAACCTATGTCAAGCAGG 523
QY 521 TGGTGGGCAAGCTGGGCTGCTGAGTACAGCAGCATGGAGTCAGTCCCTCGGGCTGCC 580
DB 524 AGGCTGAAGGATTACATGCTGAGTATCCAAACTTGAAGAACTAAATCTCTCCCT 583
QY 581 TCATGCTCTGGTGGCTTGGCTTGGCTTCTGTGGGCGAGTGGGATCATCTCTCTCT 640
DB 584 GGATTTCTGTGGGCGATGTTTCATAGGATATGATCTTCACTTAATCATCTTCATCT 643
QY 641 GCTATATGAAGATCACCTGGAAGCTGTGAGCAGCATGCTGGGAGGACCCAGTCAGCAGCA 700
DB 644 GCTATTTCTAGATCTGTGTGAACCTTCTCAGACTGCCAAACCAACCACTCACTGAGA 703
QY 701 GGAAGAGACCCACTGCGCAGGCTGCTGCTTACGCTGTGATGCTGTGGCGGTGGTG 760

DB 704 AATCTGGTAAACAAAAAGGCTCTCAACACATATTCTTATTATTGTTGTTGTTTC 763
QY 761 TCTGCTTACGCCCTTACACCTCAACATCAAGCAGTTCTATGGCAGAGGATGCTCCACC 820
DB 764 TCTGTTTCAACCTTACCATGTTTGAATATTCAACATATGATTAAAGAGCTTCGTTCT 823
QY 821 TGCCATCC-----TGTGGCAGCGAGGCTTCTTACTGTCCCTTCAGGCCACCGTGG 874
DB 824 CTAATTTCTGGTAATGAGCAAGACATTCGTTCCAGATTCTCTGCATTTACAGTAT 883
QY 875 CCTCATGAACATGAATCTGTCGATTTACCCCAATCAATTTACTTCTTTGCAATCCCACT 934
DB 884 GCCTGATGAATTTCAATTTGTCATGACCCCTTTTATCTACTTCTTTGCAATGTAAGGGT 943
QY 935 ACAGGAATGCTCTGCGGATTTTAAAGCTCAAGGGTCTTCTCTCTCTCTCTCTCTCTCT 994
DB 944 ATAAGGAAGGTTATGAGATCTGTAAGCGCACTGATGATGATTTCTAGTGTCTGCTG 1003
QY 995 CCTCTCTCCACCCAGGAAAGCTTCTTCAAGAACACA 1032
DB 1004 TGAAGTCAGCCCTCGAAGAAATTCAGTGAATGACA 1041
RESULT 11
ACC48996
ID ACC48996 standard; cDNA; 1086 BP.
XX
AC ACC48996;
XX
DT 11-AUG-2003 (first entry)
XX
DE Human G-protein coupled receptor EBI2 coding sequence.
XX
KW Human; EBI2; G-protein coupled receptor; receptor; neuropeptide;
KW corticotropin releasing factor; antidepressant; tranquilizer; nootropic;
KW neuroprotective; cerebroprotective; antiinflammatory; antirheumatic;
KW antiarthritic; anorectic; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1086
FT /tag= a
FT /product= "Human EBI2"
XX
FN EPI286165-A2.
XX
PD 26-FEB-2003.
XX
PE 01-AUG-2002; 2002EP-00255380.
XX
PR 15-AUG-2001; 2001GB-00019920.
XX
PA (PFIZ) PFIZER LTD.
PA (PFIZ) PFIZER INC.
XX
PI Fidock MD, Robas NW;
XX
XX WPI; 2003-344719/33.
DR P-PSDB; ABR42051.
PT
PT Use of urotensin I or sauvagine, or their analogue or mimetic, as ligand
PT or modulator of EBI2 or for diagnosing or treating depression, anxiety-
PT related disorders, stress, Alzheimer's disease, stroke, inflammation or
PT obesity.
XX
PS Claim 8; Page 22; 29pp; English.
XX
CC The present sequence is the coding sequence for human G-protein coupled
CC receptor EBI2. The invention relates to the identification of EBI2 as a
CC receptor of neuropeptides of the corticotropin releasing factor (CRF)
CC family. These neuropeptides, e.g. urotensin I or sauvagine (and their
CC analogues or mimetics), can be used as modulators of EBI2. The invention

CC also provides screening methods to identify agonists and antagonists for
CC EBI2. Modulators of the receptor may find use in diagnosing or treating
CC depression, anxiety-related disorders (e.g. panic), stress, Alzheimer's
CC disease, stroke, inflammatory disorders (e.g. rheumatoid arthritis),
CC eating disorders (e.g. anorexia) and obesity

XX Sequence 1086 BP; 314 A; 242 C; 206 G; 324 T; 0 U; 0 Other;

Query Match 16.7%; Score 184.8; DB 7; Length 1086;
Best Local Similarity 50.6%; Pred. No. 4.1e-34;
Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;

QY 101 CTCTGTTCTACAGCCCTCTCTGCTCTTCACTGCTCCCTGGGAAACATCTTGGCTTTGCC 160
Db 104 CTCTGCAATACAGCTCTCTCTTCACTATGCTGCTGGGAAACTTACTAGCTTGGTGG 163
QY 161 TTACTGTCAAGAGCAGGAGAGATCAACTGCACAGCATCTACTCTGGTGGACCTGGCTG 220
Db 164 TCATTGTTCAAAACAGGAAATAAATCACTTACCCTCTTATCAACAAATTTGGTGA 223
QY 221 TGTCTGACCTCTCTTCCACGTGGCTTACCGGAAAGGTGGTGTGTATGTGCTGGCT 280
Db 224 TTTCTGATATATCTTTTACACGGCTTTGCTACAGATAGCTTACTATGCAATGGGCT 283
QY 281 CCAGCTGGCTTTCCGCAAGGGCTCTGAGGCTGAGGGCTTGTGCTCTACACCGCA 340
Db 284 TTGACTGAGAAATCGGAGATCCCTTGTGTAGGATACTGGCTAGTGTTTTACATCAAC 343
QY 341 CCTAGGGGGGCTCTACTCTCATGSCCTGTGTGAGCGTGGACCATTACCAGCTGTGGCT 400
Db 344 CATATGCAAGTGTGAACCTTTATGACCTGCTGAGTATGACGCTTCAITGCTGTGGTGC 403
QY 401 GTGCCACTGGGGCCCGCTCCGACGGCTGGCGCCGACGCTGCTGGTGGCA 460
Db 404 ACCCTTACGTACACAAAGNAAAGAGATGAACATGCAAAAGCGGTGTGATATTG 463
QY 461 TCTGACCTTGGTCTGCTGCAGACGATGCCCTTGTCTTTGATGCCCATGACCAAGCCG 520
Db 464 TCTGATCTTGTAGTATTGTCTCAGACACTCCCTCTCATCAACCTCTATGTCAAAGCAG 523
QY 521 TGGTGGGCAAGCTGGCTGATGAGTACAGCAGCATGAGTCACTCGGGCTGCC 580
Db 524 AGGCTGAAGATACATGATGAGTATCCAACTTTGAGAACTAATCTCTCCCT 583
QY 581 TCATGGTCTGGTGGCTTTGCCATGGCTCTGTGGCCAGTGGGGATCATCTGTCT 640
Db 584 GGAATCTGCTTGGGCACTGTTTCAAGATATGATCTTCCATTAATCAATCTCTCATCT 643
QY 641 GCTATATGAGATCACCTGGAAGCTGTGCAGCAGCTCGGAGGACCCAGTGAACAGCA 700
Db 644 GCTATCTCAGATCTGTGCAACTCTTCAAGATGCAAAACCAACCACTCACTGAGA 703
QY 701 GGAAGGACGCCACTGGCGAGGCTGCTGTGCTTACGCTGTGCTGGTGGCGGTGG 760
Db 704 AATCTGGTGTAAACAAAGGCTCTCAACCAATTAATCTTATTTGTTGTTGTTTC 763
QY 761 TCTGTTCTCAGCCCTTACCACCTCAACATCAAGCATGTTTCATGGGAGAGGATGCTCCAC 820
Db 764 TCTGTTTCAACCTTACATCTTGCATATTTCACATATGATGAGAAGCTTGGTTCT 823
QY 821 TGCCATCC-----TGTGCGAGCGGAGGGCTTTCTTACTGTGCTTCAAGCCACCGTGG 874
Db 824 CTAATTTCTCTGGAATGTAGCCAAAGACATTCGTTTCCAGATTTCTCTGCACTTTACAGTAT 883
QY 875 CCCTCATCAACATGAATCTGGGCAATTACCCCAATCTTTTACTTCTTTGATCCACCATTT 934
Db 884 GCTTATGATGAATTCATTTGCTGATGAGACCTTTTATCTACTTCTTTGATGTAAGGTT 943
QY 935 ACAGAAATGCTCTGGGCATTTTAAAGCTCAAGGGTCTTCTCTCTCTCTCTCTCTCTCT 994
Db 944 ATAAGAGAAAGGTTATGAGGATGCTGAAACGGCAAGTCAGTGTATGATGATTTCTAGTGTG 1003
QY 995 COTCTCCACCCAGGAAAGCTTCTTTCAGAAACACCA 1032

RESULT 12

AD22596
ID ADC22596 standard; cDNA; 1086 BP.

XX ADC22596;

AC ADC22596;

XX 18-DEC-2003 (first entry)

XX Human G protein-coupled receptor cDNA #22.

XX Human; gene; ss; G protein-coupled receptor; GPCR;

XX transmembrane-6 region; TM6; intracellular-3 region; IC3.

XX Homo sapiens.

XX US655339-B1.

XX 29-APR-2003.

XX 13-OCT-1998; 98US-00170496.

XX 14-APR-1997; 97US-00839449.

XX 14-APR-1998; 98US-00060188.

XX 26-JUN-1998; 98US-0090783P.

XX 07-AUG-1998; 98US-0095677P.

XX (AREN-) ARENA PHARM INC.

XX Liaw CW, Behan DP, Chalmers DT;

XX WPI; 2003-742861/70.

XX P-PSDB; ADC22597.

XX Creating a constitutively active version of an endogenous human G protein

XX coupled receptor (GPCR) comprises substituting a specific amino acid in

XX the transmembrane-6 region with a different amino acid, and testing for

XX constitutive activity.

XX Example 1; SEQ ID NO 77; 221pp; English.

XX The invention relates to a method for treating a non-endogenous,

XX constitutively active version of an endogenous human G protein-coupled

XX receptor (GPCR) that has a transmembrane-6 (TM6) region and an

XX intracellular-3 (IC3) region, by substituting a specific amino acid in

XX the TM6 region with a different amino acid, and testing for constitutive

XX activity. The method is useful for creating a constitutively active

XX version of an endogenous human GPCR that comprises a transmembrane 6

XX region and an intracellular loop 3 region. The altered human GPCR

XX polypeptides are useful for screening test compounds for identification

XX of inverse agonists or partial agonists of GPCR polypeptides, which may

XX have therapeutic uses. The altered GPCRs may also be used in vivo or in

XX vitro in biological research. A nucleic acid encoding the altered GPCR

XX may be used to create a transgenic animal expressing the altered GPCR.

XX The method allows screening for compounds that modulate the activity of a

XX human G protein-coupled receptor without the need for provision of a

XX ligand for the receptor. This is particularly useful in allowing

XX screening of compounds against orphan receptors for which no ligand is

XX currently known. This sequence represents cDNA encoding a human GPCR

XX polypeptide of the invention.

XX Sequence 1086 BP; 314 A; 242 C; 206 G; 324 T; 0 U; 0 Other;

QY Query Match

Db Best Local Similarity

16.7%; Score 184.8; DB 9; Length 1086;

50.6%; Pred. No. 4.1e-34;

Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;

QY 101 CTCTGTTCTACAGCCCTCTCTGCTCTTCACTGCTCCCTGGGAAACATCTTGGCTTTGCC 160

Db 104 CTCTGCAATACAGCTCTCTCTTCACTATGCTGCTGGGAAACTTACTAGCTTGGTGG 163

161 TTACTGTCAAAAGAGCAGGAGATCAACTGCACAGGCATCTACTGCTGCACCTGGCTG 220
164 TCATTGTTCAAAACGAGAAATAATCAACTTACCACTCTATTCACAAATTTGGTGA 223
221 TGTCTGACCTGCTGTTACCGTGGCTTTACCGGAGAGGTTGGTGTATGTATGTGCTGGCT 280
224 TTTCTGATATATCTTTTACCACGGCTTTGGCTTACACGAATAGCTTACTATGCAATGGCT 283
281 CCAGCTGGCTTTTGGCAAGGGCTCTGACGGCTGACGGCTTTGTGCTCTACACCGACA 340
284 TTGACTGGAGATCGAGATGCTTGTGTAGATACCTGCCCTAGTGTATTCATACACA 343
341 COTACGGGGGCTTACCTATGCTGTGTAGCTGTGAGCTGGAACATACCCAGCTGTGTCT 400
344 CATATGCAGGTGTGAATTTATGACCTGCCTGAGTATTTGACCGCTTCATTTCTGTGTG 403
401 GTGCCACTGGGGCCGGCTCCGACAGCTGGCGCGCAGGCTGTCTGCGTGGCCA 460
404 ACCCTTACGCTAACCAAGATATAAAGGATTTGAACATGCAAAAGGCTGTGCAATTTG 463
461 TCTGACCTTGTGCTGTGCTGAGAGATGCGCTTCTTGTATGCCCATGACCAAGCGC 520
464 TCTGATTTCTAGTATTTGCTGACAGACTCCCACTCTCATCAACCTATGTCAAAAGCAG 523
521 TGTGGGCAAGCTGCGCTGCATGAGATGACAGAGATGAGTCACTCTCGGGCTGCC 580
524 AGGCTGAAGGATATCATGCTGAGTATCCAACTTTGAAGAACTAAATCTCTTCCCT 583
581 TCATGCTCTGTGGCTTTGCTATGCTTCTGTGGGCTGAGTGGGATCATCTGTCTCT 640
584 GGAATCTGCTGGGCACTTTTCATAGGATATGTTCCACTATATCATCTCTCATCT 643
641 GGTATGAGATACCTGGAAGCTGTGAGCAGAGCTCGGAGGAGCCAGTGCACGACA 700
644 GCTATTTCTGATCTGCTGCAACTCTTCAAGCTGCAAAACCAAAACCACTCACTGAGA 703
701 GGAAGGAGCGCACTGGCGAGCTGCTTACGCTGTGATGCTGTGCTGGCGCTGGGG 760
704 AATCTGGTGAACAAAGGCTCTCAACAAATTTATTTATTTGTTGTTGTTGTTG 763
761 TCTGCTTACGCCCCCTACCACTCAATCAAGAGTTCATGCGGAGAGGATGCTCCACC 820
764 TCTGTTTACACCTTACCATGTTGCAATTTATTCACATATGATTAAGAGAGCTTCTTCT 823
821 TGCCATCC-----TGTGCGAGCGGAGGCTTCTTACTGTCCTTCAGGCCACCGTGG 874
824 CTAATTTCTGGAATGTAGCCAAAGACATTCGTTCCAGATTTCTCTGCACCTTACGAT 883
875 CCTCATGAACATGAACCTGTGGCATTTACCCCAATCATTTTACTTCTTTCATCCACCA 934
884 GCTGATGAATTCATTTCTGATGAGCCCTTTTATCTACTTCTTTCATGTAAGGGT 943
935 AAGGAAATGGCTTCCTGGGCAATTTAAGCTCAAGGGTCTTCTCTCTCTCTCTCTCTCT 994
944 ATAAGAGAAGGTATGAGGATGCTGAAACGGCAAGTCAAGTATCGATTTCTAGTGTG 1003
995 CTTCTCCACCCCGAGAAAGCTTCTTCAGAAACACCA 1032
1004 TGAAGTCAGCCCTGAGAAATTCACGTGAATGACA 1041

RESULT 13

AAQ64126

ID AAQ64126 standard; DNA; 1643 BP.

XX AAQ64126;

XX AC

XX 25-MAR-2003 (revised)

DT 03-FEB-1995 (first entry)

XX

DE Epstein Barr virus induced (EBI-2) gene.

XX

KW Epstein Barr virus; EBV; induction; detection; diagnosis; lymphocytes;
antigen; growth; differentiation; mediator; infectious mononucleosis; ss.
OS Homo sapiens.

XX Key Location/Qualifiers
FH 34..1119
FT /tag= a
FT /product= "Epstein Barr virus induced polypeptide."

PN WO9412519-A1.

XX 09-JUN-1994.

XX 08-OCT-1993; 93WO-US009636.

XX 25-NOV-1992; 92US-00980518.

XX (BGM) BRIGHAM & WOMENS HOSPITAL.

XX Birkenbach M, Kieff E;

XX WPI; 1994-200183/24.

XX P-PSDB; AAR54080.

PT DNA coding for Epstein Barr Virus induced (EBI) polypeptide(s) and
antibodies to EBI1, 2 and 3 - useful for detecting EBV by hybridisation
or by immunocassay.

XX Claim 2; Page 56-58; 84pp; English.

XX EBV infected B lymphocytes recapitulate features of antigen stimulation
in enlarging, increasing RNA synthesis, expressing activation antigens
and adhesion molecules, secreting Ig and proliferating. Unlike antigen
stimulated B lymphocytes, EBV infected B lymphocytes continue to
proliferate (in vitro) as immortalised lymphoblastoid cell lines. Because
of the similar effects of EBV and antigen, EBV induced genes are likely
to include mediators of antigen induced B lymphocyte growth or
differentiation. (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 1643 BP; 509 A; 339 C; 285 G; 510 T; 0 U; 0 Other;

Query Match 16.7%; Score 184.8; DB 2; Length 1643;

Best Local Similarity 50.6%; Pred. No. 4.8e-34;

Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;

QY 101 CTCTGTTTACACAGCCCTCTGCTTTCAGTCCCTCGGAAACATCTTGCCTTTGCC 160

Db 137 CTCTGCAATTACGCCCTGCTTTCATCTGGGCTCGTGGAACTTACTAGCCTTGGTCG 196

QY 161 TTACCTGTCAAAAGAGCAGGAAGATCAACTGCAACAGGCATCTTACCTGGTGCACTGGCTG 220

Db 197 TCATTGTTCAAAACAGGAAAAAATCAACTCTACCCCTCTATTCAACAAATTTGGTGA 256

QY 221 TGTCTGACCTGCTTTCACCGTGGCTTACCGGAAAGGGTGGTGTGTATGTGCTGGCT 280

Db 257 TTTCTGATATACTTTTTACCACGGCTTTGCTTACAGATAGCTACTATGCAATGGCT 316

QY 281 CCAGCTGGCCTTTTCGCAAGGGCTCTGCAAGGCTGACGGCTTTGTCTTACCCGACA 340

Db 317 TTGACTGAGAAATCGGAGATGCTTGTGTAGGATAACTGGCTAGTGTATTACATCAACA 376

QY 341 CTTACGGGGGGTCTTACCTCATGGCTGTGTAGCTGGACCATTACCCAGCTGTGTCT 400

Db 377 CATATGCAAGTGTGAACCTTTATGACCTGCTGTAGTATTGACCGCTTCATTGTGTGTG 436

QY 401 GTGCCACTGGGGCCCGGCTCCGACGGCTGGCGCGCAGGCTGGTCTGGTGGGCCA 460

Db 437 ACCCTCTACGCTACACAAAGATTAAGAGATTGAACATGCAAAAGGCTGTGCATATTG 496

QY 461 TCTGGACCTTGGTCTGCTGCAAGATGCTTGTCTTGTATGCCCATGACCAAGCCGC 520

Db 497 TCTGGATTCTAGTATTTGCTCAGACACTCCCACTCCCTCATCAACCCCTATGTCAAAGCAG 556

QY 521 TGGTGGCAAGTGGCCCTGATGGAGTACAGCAGATGAGTCTCGGCTCCCC 580
 Db 557 AGGCTGAAGAGTATCATGATGAGTATCCAAACTTTGAAGAACTTATCTTCCCT 616
 QY 581 TCATGGTCTGCTGGCCCTTGGCCATTTGCTGCGCCAGTGGGATCATCTGTCCT 640
 Db 617 GGAATCTGCTGGGCGATGTTTATAGATATGATCTTCCATTTATATCTCATCT 676
 QY 641 GCTATATGAAGATCACTGGAAGTGTGCAGCAGCTGGGAGGACCCAGTACACAGA 700
 Db 677 GCTATTTCTGATCTGCTGCAAACTCTTCAAGACTGCAAAACCAAACTCACTGAGA 736
 QY 701 GGAAGGACGCCACTGGGAGGCTGCTGCTTACGCTGCTGCTGCTGCTGCTGCTG 760
 Db 737 AATCTGGTGTAAACAAAAGGCTCTCAACAAATTTCTTATTTGTTGTTGTTTC 796
 QY 761 TCTGCTTACGCCCTTACCACTCAACATCAAGAGTGTATGCGGAGGAGTCTCCACC 820
 Db 797 TCTGTTTACACCTTACCATGTTGCAATTTATCAACATATGATTAAGAAGCTTCTGTTCT 856
 QY 821 TGCCATCC-----TGTGCCAGCGGAGGCTTTCTTACTGTCTGCTTACGCGCACCGTGG 874
 Db 857 CTAAATTTCTGGAATGTAGCAAGACATTCGTTCCAGATTTCTCTGCATTTACAGTAT 916
 QY 875 CCTCATGAACATGACTGTGGATATACCCCAATCATTTACTTTTGGATCCACCATT 934
 Db 917 GCCTCATGAATCAATTTGCTGCTGAGACCTTTTATCTACTTTTGGCATGTAAGAGGT 976
 QY 935 ACAGAAATGCTCTCGGCGATTTTAAAGCTCAAGAGGTCTTCTCTCTCTCTCTCTCT 994
 Db 977 ATACAGAAAGTTATAGGATGCTGAACCGCAAGTCAGTGTATCGATTTCTAGTCTG 1036
 QY 995 CCTCTCCACCCAGGAAAGCTTCTTCAGAAACACCA 1032
 Db 1037 TGAAGTACGCCCTTGAGAAATTCACGTGAATGCA 1074

RESULT 14

AAV25491
 ID AAV25491 standard; cDNA; 1643 BP.
 XX AAV25491;
 AC AAV25491;
 XX
 DT 09-JUL-1998 (first entry)
 XX
 DE cDNA for Epstein Barr virus induced gene 2 (EBI-2).
 XX
 KW Assessing; monitoring; foetal development; placental development;
 KW Epstein Barr virus; EBV; induced gene 2; EBI-2; ds.
 XX

Homo sapiens.

XX
 XX Key Location/Qualifiers
 FT CDS 34..1119
 FT /*tag= a

US5744301-A.

28-APR-1998.

XX 02-FEB-1995; 95US-00383750.

XX 25-NOV-1992; 92US-00980518.

XX 30-NOV-1994; 94US-00352678.

XX (BGM) BRIGHAM & WOMENS HOSPITAL.

XX Kieff E, Birkenbach M;

XX WPI; 1998-271060/24.

XX P-PSDB; AAW53623.

XX

PT Assessing or monitoring foetal or placental development - comprises
 PT detecting the level or size of Epstein Barr virus induced nucleic acid or
 XX protein in maternal serum samples.

PS Example 2; Col 43-46; 45pp; English.

XX The present sequence was used in the development of a novel method for
 CC assessing or monitoring foetal or placental development. The method
 CC comprises taking a maternal serum sample, and detecting the level or size
 CC of Epstein Barr virus (EBV) induced gene or protein 3 (EBI-3) to obtain a
 CC result, which can be compared to a control to assess or monitor foetal or
 CC placental development

XX Sequence 1643 BP; 509 A; 339 C; 285 G; 510 T; 0 U; 0 Other;

Query March 16.78; Score 184.8; DB 2; Length 1643;

Best Local Similarity 50.6%; Pred. No. 4.9e-34;

Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;

QY 101 CTCGTGTTCTACACAGCCCTCTGCTGTTTCAGTCCCTGGGAAACATCTTGGCCCTTGGC 160

Db 137 CTCTGCATTACAGCCTCGTCTTCATCATTTGGGCTCGTGGGAACTTACTAGCCTTGGTCG 196

QY 161 TTACTCTCAAAAGACAGGAGATCACTGCACAGGATCTTACCTGTCGACCTGGCTG 220

Db 197 TCATTTGTTCAAAACAGGAAAAAATCACTTACCACCTCTATTCAACAAATTTGGTGA 256

QY 221 TGTCTGACCTGCTGTTTCCCGTGGCTTACCGGAAAGGCTGTTGTTATGCTGGGCT 280

Db 257 TTTCTGATATACATTTTACCAAGCTTGGCTACACGAATAGCTACTATGCAATGGGCT 316

QY 281 CCAGCTGGCTTTGGCAAGGGCTCTGCAGGCTGACGGCTTTGCTCTACACCGCA 340

Db 317 TTGACTGGAGAAATCGGAGATGCTTGTAGAGATACTCGCTAGTGTATACATCAACA 376

QY 341 CCTACGGGGGCTTACCTCATGGCTGTGTGAGCTGGACCATTACCAGCTGTGGCT 400

Db 377 CATATGCAAGTGTGAACCTTTATGACCTGCTGAGTATTGACCGCTTCATTGCTGTGGTC 436

QY 401 GTGCCACTGGGCGCCGCTCCGACGCTGGCGCGCGCGCGCGCGCTGCTCGTGGCCA 460

Db 437 ACCCTTACCTACAACAGATAAAAGGATTAACATGAAGGCGTGTGATATTG 496

QY 461 TCTGGACCTTTGGTGTCTGCTGCAGACGATGCCCTTGTCTTGTATGCCATGACCAAGCGC 520

Db 497 TCTGGATTTCTAGTATTGCTTCAGACACTCCCACTCCTCATCAACCTATGTCAAGACAG 556

QY 521 TGTGGSCAAGCTGGCTGCTGATGAGTACAGCAGATGAGTCACTCCTCGGCTGCCCC 580

Db 557 AGGCTGAAGATTAATGATGAGGATATCCAACTTTGAAGAACTAAATCTCTTCCCT 616

QY 581 TCATGGTCTGTGGCTTTGCCATTTGGCTTTGTGGGCCAGTGGGATCATCTGTCT 640

Db 617 GGATTTCTGCTTGGGCGATGTTTCATAGGATATGTACTTCCACTTATTAATCATTTCTCATCT 676

QY 641 GCTATATGAAGATCACTGGAAGCTGTGCAGCAGCTCGGAGGACCCAGTACACAGA 700

Db 677 GCTATTTCTGATCTGCTGCAAACTTTTCAAGTCTGCAAAACCAACCACTCATCTGAGA 736

QY 701 GGAAGGACGCCACTGGGAGGCTGCTGCTTACGCTGCTGATGCTGGGCGCTGGTGG 760

Db 737 AATCTGGTGTAAACAAAAGGCTCTCAACAAATTTCTTATTTATTTGTTGTTGTTTC 796

QY 761 TCTGCTTACGCCCTTACCACTCAACATCAAGAGTTCATGCGGAGAGGATGCTCCACC 820

Db 797 TCTGTTTACACCTTACCATTTGCAATTTATCAACATATGATTAAGAAGCTTCGTTCT 856

QY 821 TGCCATCC-----TGTGCCAGCGGAGGCTTTCTTACTGCTCCCTTACGGCCACCGTGG 874

Db 857 CTAATTTCTTGGAAATGTAGCAAAAGCATTCGTTCCAGATTTCTCTGCACTTTACAGTAT 916

QY 875 CCCTCATGAACATGAATGTGGCATTACCCCAATCATTTACTTCTTGTGATCCACCATTT 934

Db 917 GCCTGATGAACCTCAATTGCTGCATGACCCCTTTTATCTACTCTTTTGCATGTAAGGGT 976
QY 935 ACAGGAATGCTCCTGGGCAATTTAAAGCTCAAGGGTCTTCTCTCTCTCTCTCTCTCTCT 994
Db 977 ATAAGAGAAGGTTATGAGGATGCTGAAACGGCAAGTCAGTGATCGATCTTCTAGTCTG 1036
QY 995 CCTCTCCACCCAGGAAGGCTTCTTCAGAAACACCA 1032
Db 1037 TGAAGTCAGCCCTGAGAAATTCAGTGAATGACA 1074

RESULT 15

ACAS6681
ID ACAS6681 standard; cDNA; 1643 BP.
XX ACAS6681;
XX AC
XX AC
XX DT
XX 06-JUN-2003 (first entry)
XX Human signalling pathway polynucleotide probe SEQ ID NO 1279.
DE Human; probe; ss; array element; Parkinson's disease;
XX Human; probe; ss; array element; Parkinson's disease;
KW signalling pathway population; cancer; adenocarcinoma; leukaemia;
KW immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.
XX Homo sapiens.
OS
XX
XX US6500938-B1.
XX 31-DEC-2002.
XX 30-JAN-1998; 98US-00016434.
XX 30-JAN-1998; 98US-00016434.
XX (INV-) INCYTE GENOMICS INC.
XX Au-Young J, Seilhamer JJ;
XX WPI; 2003-352189/33.
XX Combination of polynucleotide probes, useful as array elements in a
PT microarray for monitoring the expression of a number of target
PT polynucleotides.
XX Claim 1; SEQ ID NO 1279; 65pp; English.

XX The invention relates to a combination which, comprises a number of
XX polynucleotide probes comprising a sequence selected from one of the 1490
XX sequences mentioned in the specification. The combination is useful as an
XX array element in a microarray for monitoring the expression of a number
XX of target polynucleotides. The microarray is particularly useful in the
XX diagnosis and treatment of cancer and immunopathology and neuropathology.
XX The microarray is useful in diagnostics and treatment regimens, drug
XX discovery and development, toxicological and carcinogenicity studies,
XX forensics and pharmacogenomics. The microarray is also useful for
XX monitoring progression of diseases and for developing sophisticated
XX profiles for the effects of currently available therapeutic drugs. The
XX combination is also useful for purifying a subpopulation of mRNAs, cDNAs
XX and genomic fragments and in research and diagnostic applications. The
XX array can detect changes in expression in a large number of genes coding
XX for different signalling pathway populations which can be used to diagnose
XX various diseases including cancer e.g. adenocarcinoma and leukaemia,
XX immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease
XX and Parkinson's disease. The present sequence represents a polynucleotide
XX probe of the invention. Note: The sequence data for this patent did not
XX form part of the printed specification but was obtained in electronic
XX format directly from USPTO at
XX seqdata.uspto.gov/sequence.html?DocID=06500938B1

XX Sequence 1643 BP; 509 A; 339 C; 285 G; 510 T; 0 U; 0 Other;

XX Query Match 16.7%; Score 184.8; DB 7; Length 1643;

Best Local Similarity 50.6%; Pred. No. 4.8e-34;
Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;
QY 101 CRTCTGTTCTACACAGCCCTCTCGGCTCTTCAAGTCCCTGGGAAACATCCTTGCCTTTGCC 160
Db 137 CTTCTGCAATACAGCCCTCGTCTTCATCATTTGGGCTCGTGGGAACTTACTAGACCTTTGGTGG 196
QY 161 TTACCTGTCAAAAGAGCAGGAGATCAACTGCACAGGATCTACCTGGTGCACCTGGGTG 220
Db 197 TCATTGTTCAAAACAGGAAAAAATCAACTCTACACCTCTATTTCACAAATTTGGTGA 256
QY 221 TGTCTGACCTGCTGTTTACCCGCTGGGCTTACCGGGAAGGTTGGTGTATTGCTGGGCT 280
Db 257 TTTCTGATATACCTTTTACACGGCTTTGCTTACAGGATACCTTACTATGCAATGGGCT 316
QY 281 CCAGCTGGGCTTTGGCAAGGGGCTCTGCAGGCTGACGGCTTTGTGCTCTACACGACA 340
Db 317 TTGACTGGAGAAATCGGAGATGCTTGTGTAGGATAACTGCGCTAGTGTATTTTACATCAACA 376
QY 341 CCTACGGGGGGTCTACCTCATGGGCTGTGTGAGGCTGGACCATTACCAGCTGTGGTCT 400
Db 377 CATATGACAGTGTGAACCTTTATGACCTGCTGCTGATTTGACCGCTTCATTGTGTGGTGC 436
QY 401 GTGCCCACTGGGGCCCGCCCTCCGCAAGGCTGGCGGCGCCAGGCTGTGCTGGTGGCCA 460
Db 437 ACCCTCTACGCTACAAACAGATAAAAGGATTAACATGCAAAAGCGGTGTGCATATTG 496
QY 461 TCTGGACCTTTGGTGTGCTGCAGACGATGCTTGTCTTGTGATGCCCATGACCAAGCCGC 520
Db 497 TCTGGATTTCTAGTATTGCTGCTCAGACACTCCACTCCTCATCAACCTATGTCAAAGCAGG 556
QY 521 TGGTGGGCAAGCTGGCTGCTATGGAGTACAGGAGGATGGAGTCACTCTGGGCTGCCCC 580
Db 557 AGGCTGAAAGGATTACATGCTGGAGTATCCAAACTTTTGAAGAACTAAATCTCTTCCCT 616
QY 581 TCATGCTCTGTGGGCTTTTGGCAATTTGCTTGTGGGCGAGTGGGATCATCTCTCTCT 640
Db 617 GGATTTCTGCTTGGGCGATGTTTCTAGGATATGTACTTCACTTATATATCTCTCATCT 676
QY 641 GCTATATGAAGATCACTGTGAAGCTGTGACGACAGCTCGGAGGAGCCAGTGAACAGCA 700
Db 677 GCTATTCTCAGATCTGCTGCAAACTCTTCAGAACTGCCAAACAAACCCACTCACTGAGA 736
QY 701 GGAAGGACGCCACTGGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 760
Db 737 AATCTGCTGTAAACAAAGGCTCTCAACACAAATTTCTTATTTGTTGTTGTTTCT 796
QY 761 TCTGCTTCAGCCCTTACCACCTCAACATCAAGCAGTTTCATGGCGAGAGGATCTCCACC 820
Db 797 TCTGTTTACACCTTACCCTTACCTGTTGCAATTTTCAACATATGATTAAAGAGCTTCTCT 856
QY 821 TGCCATGCC-----TGTGGCGAGGAGGCTTTCTTACTGTCTCTTCCAGCCACCGTGG 874
Db 857 CTAATTTCTCTGGAATGTAGCCAAAGACATTCGTTCCAGATTTCTCTGCACTTTACAGTAT 916
QY 875 CCCTCATGAACATGAAGTGTGGCATTAACCCCAATCAATTTACTTTTGTGATCCACCCATT 934
Db 917 GCCTGATGAATCTCAATTTGCTGATGGACCTTTTATCTACTTTTGTGATGTAAGGGT 976
QY 935 ACAGGAATGCTCTGGGCAATTTAAAGCTCAAGGCTTCTCTCTCTCTCTCTCTCTCTCTCT 994
Db 977 ATAAGAAAGGTTATGAGGATGCTGAAACGGCAAGTCAGTGTATCGATTTCTAGTCTG 1036
QY 995 CCTCTCCACCCAGGAAAAAGCTTCTTCAGAAACACCA 1032
Db 1037 TGAAGTCAGCCCTGAGAAATTCAGTGAATGACA 1074

Search completed: February 21, 2004, 19:27:06
Job time : 494 secs

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	184.8	16.7	1086	4	US-09-170-496D-77	Sequence 77, Appl
2	184.8	16.7	1643	1	US-08-383-750-3	Sequence 3, Appl
3	184.8	16.7	1643	3	US-08-352-678-3	Sequence 3, Appl
4	184.8	16.7	1643	4	US-09-536-954-3	Sequence 3, Appl
5	184.8	16.7	1643	4	US-09-016-434-1279	Sequence 1279, Ap
6	184.8	16.7	1643	5	PCF-U893-09363-6	Sequence 3, Appl
7	183.2	16.6	1086	4	US-09-170-496D-205	Sequence 205, App
8	118.6	10.7	2025	4	US-09-016-434-1482	Sequence 1482, Ap
9	112	10.1	1020	4	US-09-170-496D-31	Sequence 31, Appl
10	112	10.1	1020	4	US-09-170-496D-181	Sequence 181, App
11	112	10.1	1900	4	US-09-016-434-1484	Sequence 1484, Ap
12	112	10.1	1901	1	US-08-153-848-43	Sequence 43, Appl
13	112	10.1	1901	3	US-09-289-843A-43	Sequence 43, Appl
14	112	10.1	1901	4	US-09-088-337B-43	Sequence 43, Appl
15	112	10.1	1901	5	PCF-U893-11153-43	Sequence 43, Appl
16	112	10.1	2453	5	PCF-U893-07180-1	Sequence 1, Appl
17	106.8	9.7	2218	4	US-09-214-904-3	Sequence 3, Appl
18	106.8	9.7	2219	4	US-08-432-174A-1	Sequence 1, Appl
19	106.8	9.7	2272	3	US-08-147-592A-3	Sequence 3, Appl
20	106.8	9.7	2272	4	US-08-292-694A-3	Sequence 3, Appl
21	105.4	9.5	1179	4	US-08-148-708-3	Sequence 3, Appl
22	105.4	9.5	1378	1	US-08-753-848-2	Sequence 2, Appl
23	105.4	9.5	1378	4	US-08-148-708-6	Sequence 6, Appl
24	105.4	9.5	1378	4	US-09-016-434-1123	Sequence 1123, Ap
25	105.4	9.5	1378	5	PCF-U895-09383-2	Sequence 2, Appl
26	105.2	9.5	1829	2	US-08-411-859-1	Sequence 1, Appl
27	105.2	9.5	1829	3	US-08-387-707-7	Sequence 7, Appl

Db 464 TCTGATCTAGTATTGCTCAGACATCCCACTCCCTCATCAACCCCTATGCAAGCAGG 523
QY 521 TGTGGGCAAGCTGGCTGCTAGTACAGCAGATGAGTCTGCTGGGCTGCCCC 580
Db 524 AGGCTGAAAGGATTACATGCTAGTATCCAACTTTGAAGAACTTAAATCTCTCCCT 583
QY 581 TCATGCTCTGCTGGCTTTGCCATGCTGCTTCTGCTGGGCTGAGTGGGATCATCTCTCT 640
Db 584 GATTTCTGCTGGGCTGCTTCTAGGATGATGCTTCCATTTATATCTCTCTCTCTCT 643
QY 641 GCTATATGAGATCACTGGAAGTGTGACAGCAGCTGGGAGGAGCCAGTCAACGACA 700
Db 644 GCTATTTCTGATCTGCTGCAACTCTTCAAGACTGCGCAACCAAAACCCACTCACTGAGA 703
QY 701 GGAAGGAGCGGCTGGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 760
Db 704 AATCTGCTGTAACAAAAGGCTCTCAACAAATATCTTATTTATTTGTTGTTGTTTC 763
QY 761 TGTGCTTGAGCCCTTACCACCTCAACATCAAGCAGTTCATGGGAGAGGATGCTCAAC 820
Db 764 TCTGTTTCAACCTTACCATGTTGCAATTAATCAACATATGATTAAGAAAGCTTCTGTTCT 823
QY 821 TGCCATCC-----TGTGCCAGGAGGCTTTCTTACTGCTGCTGCTGCTGCTGCTG 874
Db 824 CTAAATTTCTGGAATGTAGCCAAAGACATTCGTTCCAGATTTCTGCACTTTACAGTAT 883
QY 875 CCTCATGAAACATGAGTGTGGATATCCCAATCATTTACTTCTTTTGTGATCCACCAT 934
Db 884 GCTCATGAACTTCAATTTGCTGCTGAGCCCTTTTATCTACTTCTTTGCTGATGTAAGGGT 943
QY 935 ACAGAAATGCTGCTGCTGCTGCTTAAAGCTCAAGGCTGCTTCTGCTGCTGCTGCTGCT 994
Db 944 ATAAGAAAGTGTATGAGGATGCTGAAACGGCAGTCAGTGTATCGATTTCTAGTGTCTG 1003
QY 995 CCTCCTCAACCCAGGAAAGCTTCTTCAAGAACCA 1032
Db 1004 TGAAGTACGCCCTGAAGAAATTCACGTGAATGACA 1041

RESULT 2

US-08-383-750-3
; Sequence 3, Application US/08383750
; Patent No. 5744301
; GENERAL INFORMATION:
; APPLICANT: Birkenbach, Mark
; APPLICANT: Kieff, Elliot
; TITLE OF INVENTION: Epstein Barr Virus Induced Genes
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, N.W.,
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/383,750
; FILING DATE: Herewith
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Fox, Samuel, L.
; REGISTRATION NUMBER: 30,353
; REFERENCE/DOCKET NUMBER: 0827.3300001
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1643 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: both

; TOPOLOGY: linear

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 34...1116

; US-08-383-750-3

Query Match

Best Local Similarity 16.7%; Score 184.8; DB 1; Length 1643;

Matches 475; Conservative 50.6%; Pred. No. 4e-36;

Mismatches 457; Indels 6; Gaps 1;

QY 101 CTCTGTTTCTACAGCCCTCTGCTCTCAGTCCCTGGGAAACATCCTTGCCCTTTGCC 160
Db 137 CTCTGCTATACAGCTCTGCTCTTCTCATCTTGGGCTGTGGGAACTTACTAGCTTTGGTGG 196
QY 161 TTACTGCTCAAAAGAGCAGAGATCAACTGACAGGCACTTACCTGGTGACCTGGGTG 220
Db 197 TCATTGTTCAAAACAGGAAAAAATCAACTCTACCACCTCTATTTCACAAATTTGGTGA 256
QY 221 TGTCTGACCTGCTGTTTACCGTGGCTTACCGGAGGCTGTGTATTGTGTGGCT 280
Db 257 TTCTGATATCTTTTACACGGCTTTGCTTACAGATAGCTTACTATGCAATGGGCT 316
QY 281 CCAAGCTGGCCCTTTGCGGAAAGGGCTCTGCAAGCTGACGGCTTTGTCTCTACACCGACA 340
Db 317 TTGACTGGAGATCGAGATGCTTGTGTAGATTAACCTGCGCTAGTGTGTATACATCAACA 376
QY 341 CCTACGGGGGGTCTACCTCATGCTGTGTGAGGCTGGACCATACCCAGCTGTGGTCT 400
Db 377 CATATGAGCTGTGAACTTTATGACCTGCTGAGTATTGACCGCTTCATTGCTGTGGTGC 436
QY 401 GTGCCACTGGGGCCCGGCTCCGACGCTGGCGGCGGCGGCTGCTGCTGCTGCTGCTG 460
Db 437 ACCCTCTACGCTACAAAGATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 496
QY 461 TCTGGACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 520
Db 497 TCTGGATTTCTAGTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 556
QY 521 TGTGGGCAAGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 580
Db 557 AGGCTGAAAGGATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 616
QY 581 TCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 640
Db 617 GGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 676
QY 641 GCTATATGAGATCACTGGAAGCTGTGCAAGCTGCTGCAAGCTGCTGCAAGCTGCTGCA 700
Db 677 GCTATTTCTGATCTGCTGCAAGCTGCTGCAAGCTGCTGCAAGCTGCTGCAAGCTGCTG 736
QY 701 GGAAGGACGCCACTGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 760
Db 737 AATCTGCTGTAACAAAAGGCTCTCAACAAATTAATCTTATTTATTTGTTGTTGTTTC 796
QY 761 TCTGCTTCAGCCCTTACCACTCAACATCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTG 820
Db 797 TCTGTTTCAACCTTACCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 856
QY 821 TGCCATCC-----TGTGCCAGGAGGCTTTCTTACTGCTGCTGCTGCTGCTGCTGCTG 874
Db 857 CTAAATTTCTGGAATGTAGCCAAAGACATTCGTTCCAGATTTCTCTGCACTTTACAGTAT 916
QY 875 CCTCATGATGAACTGTGGCTTACCCCAATCATTTACTTCTTCTTCTGCTGCTGCTGCTG 934
Db 917 GCTGATGAACTCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 976
QY 935 ACAGGAAATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 994

Db 977 ATAGAGAAAGTTATGAGGATGCTGAAACGGCAAGTCAGTGTATCGATTTCAGTGTG 1036

Qy 995 CTTCTCCACCCAGGAAAGCTTTTCAGAAACACCA 1032
Db 1037 TGAAGTCAGCCCTGAAAGAAATTCAGTGAATGACA 1074

RESULT 3
US-08-352-678-3
; Sequence 3, Application US/08352678
; Patent No. 6043351
; GENERAL INFORMATION:
; APPLICANT: Birkenbach, Mark
; APPLICANT: Kieff, Elliott
; TITLE OF INVENTION: EPSTEIN BARR VIRUS INDUCED GENES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/352,678
; FILING DATE: 30-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/980,518
; FILING DATE: 25-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gates, Edward R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: B0801/7044
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1643 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 34...1116
; US-08-352-678-3

Query Match 16.7%; Score 184.8; DB 3; Length 1643;
Best Local Similarity 50.6%; Pred. No. 4e-36;
Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;
Qy 101 CTCCTGTTTACACGCCCTCTGCTCTTCAGTGGCCCTGGGAAACATCCTGCCCCCTTGGC 160
Db 137 CTCCTGTTTACACGCCCTCTGCTCTTCATCTGCTGCTGGGAACTTACTAGCCTTGGTCG 196
Qy 161 TTACCTGTCAAAGACGAGGAGATCACTGCACAGGCATCTACCTGGTGCACCTGGCTG 220
Db 197 TCATTGTTCAAACAGGAAATAATCACTTACCACCTCTATTCAACAATTTGGTGA 256
Qy 221 TGCTGACCTGCTGTTTACCGTGGCTTACCGGGAAGGCTGTTGTTATGCTGGCT 280
Db 257 TTTCTGATATCTTTTACACGGCTTTCCTTACACGAATAGCCTTACTATGCAATGGCT 316
Qy 281 CCAGCTGGCTTTGGCAGGGGCTCTGAGCTGACGCGCTTCTGCTCTACACCGACA 340
Db 317 TTGACTGGAGATCGAGATGCTTGTGTAGGATTAAGCTGCTAGTGTTCATCAACA 376

Qy 341 CCTACGGGGGGTCTACTCTCATGGCTGTGTGAGCTGGACCATACCCAGCTGTGGTCT 400
Db 377 CATATGCGAGGTGTGAACCTTTATGACCTCTGATGATTTGACCGCTTCTATTGCTGTGGTC 436
Qy 401 GTGCCCACTGTGGGGCCCGGCTCCGACGGCTGGCGCGCCGAGGCTGTCTGCGTGGCCA 460
Db 437 ACCCTCTACGCTACAAACAAGATTGAACATGCAAAAGCGCTGTGCATATTG 496
Qy 461 TCTGGACCTTGGTGTCTGTGACAGCATGCTTGTCTTGTATGCCCATGACCAAGCCGC 520
Db 497 TCTGGATTCTAGTATTCTGCTCAGACACTCCCACTCTCATCAACCTATGTCAAGCAGG 556
Qy 521 TGGTGGCAAGCTGGCCCTGTCATGGAGTACAGCAGCATGAGTCACTCTCGGGCTGCCCC 580
Db 557 AGGCTGAAAGGATTACATGCTGAGTATCAAACTTTGAAGAAACTAAATCTCTTCCT 616
Qy 581 TCATGCTCTGTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 640
Db 617 GGAATCTGCTTGGGCGCATGTTTCATAGGATATGTAATCTTCACTTATAATCACTCTCATCT 676
Qy 641 GCTATATGAAGATCACCTGGAAGCTGTGCAGCAGCTCGGAGGACCCAGTGCACGACA 700
Db 677 GCTATTCTCAGATCTGCTGCAAACTTTTCAGAACTTCCAAACAAACCCACTCACTGAGA 736
Qy 701 GGAAGAGCGCCACTGGGAGGCTGCTGCTTACGCTGCTGATGCTGCTGCTGCTGCTGCTG 760
Db 737 AATCTGCTGTAACAAAGGCTCTCAACACAATTTCTTATTTGTTGTTGTTGTTGTTGTTG 796
Qy 761 TCTGCTTACGCTTACCACTCAACATCAAGCAGTTCATGCGAGAGGAGTCTCCACC 820
Db 797 TCTGTTTACACCTTACCATGTTGCAATTTTCAACATATGATTAAAGAGCTTCTGTTCT 856
Qy 821 TGCCATCC-----TGTCGCGAGCGAGGGCTTTCTTACTGTCTTCTTCTGAGGCCACGGTGG 874
Db 857 CTAATTTCTGGAATGTAGCAAGACATTCGTTCCAGATTTCTCTGCACTTTTACAGTAT 916
Qy 875 CCTCATGAACATGAATCTGTCGCTTACCCCAATCAATTTACTTCTTTTGTGATCCACCAT 934
Db 917 GCTGATGAATCTCAATCTGCTGATGAGCCCTTTTACTTCTTGTGATGTAAGGCT 976
Qy 935 ACAGGAATGCTCTGCTGCGCATTTTAAAGCTCAAGGGTCTTCTCTCTCTCTCTCTCTCTCT 994
Db 977 ATAAGAGAAAGGTTATGAGGATGCTGAAACCGCAAGTCACTGTATCGATTCTTAGTGTG 1036
Qy 995 CTTCTCCACCCAGGAAAGCTTTTCAGAAACACCA 1032
Db 1037 TGAAGTCAGCCCTGAAAGAAATTCAGTGAATGACA 1074

RESULT 4
US-09-536-954-3
; Sequence 3, Application US/09536954
; Patent No. 6500926
; GENERAL INFORMATION:
; APPLICANT: Birkenbach, Mark
; APPLICANT: Kieff, Elliott
; TITLE OF INVENTION: EPSTEIN BARR VIRUS INDUCED GENES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,954

```

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/352,678
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gates, Edward R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: 50801/7044
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1643 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 34..1116
US-09-536-954-3

Query Match 16.7%; Score 184.8; DB 4; Length 1643;
Best Local Similarity 50.6%; Pred. No. 4e-36;
Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;

QY 101 CTCCTGTTCTACACAGCCCTCTGCTCTTACAGTGGGAAACATCTTGGCCCTTTGCC 160
DB 137 CTCCTGATTACAGCCCTCTGCTCTTACATCTTGGCTCGTGGGAACTTACTAGCTTGGTGC 196
QY 161 TTACTGTCAAAAGCAGGAGNATCACTGCACAGGCATCTACCTGGTGCACCTGGCTG 220
DB 197 TCATTGTTCAAAACAGGAAATACTCACTTACACCCCTCTATTCAACAAATTTGGTGA 256
QY 221 TGTCTGACCTCTCTTACCGTGGCTTACCGGAAAGGTGGTGTATTGCTTGGCT 280
DB 257 TTTCTGATATATCTTTTACACGCTTTTGGCTACAGATAGCTTACTATGCAATGGGT 316
QY 281 CCAGCTGGCTTTTCGCAAGGGCTCTGAGCTGACGGGTGGTGTCTTACACGCA 340
DB 317 TTGACTGAGAAATCGGAGATGCTTGTAGGATACTGCGCTAGTGTATTTACATCA 376
QY 341 CCTAGCGGGGGTCTACTCTCATGCGCTGTGAGCGTGGACCATATACCCAGCTGTGCT 400
DB 377 CATATGCAAGTGTGAATTTATGACCTGCGCTGATTTGACCGCTTCAITGCTGTG 436
QY 401 GTGCCACTGGGCGCGCTCGCACGCTGCGCGCGCGCTGCTGCTGCTGGCA 460
DB 437 ACCCTCTACGCTACAAAGATAAAAGGATTTGAATGCAAAAGCGGTGTGCATATTG 496
QY 461 TCTGACCTTGGTGTCTGCTGACAGATGCCCTTGTCTTGTATGCCCATGACCAAGCG 520
DB 497 TCTGATTTCTAGTATTTTGTCTCAGACACTCCCTCTCTCATCAACCTATGTCAAAG 556
QY 521 TGTGGGCAAGCTGGCTGCTGATGAGTACAGCAGCATGGAGTCAGTCTGGCTGCCCC 580
DB 557 AGGTGAAGGATTAATGATGAGATGATCAAACTTTGAAGAACTAAATCTTTCCCT 616
QY 581 TCATGGTCTGCTGGCTTTGCCATTTGGCTTCTGTGGGCCAGTGGGGATCATCTGTCT 640
DB 617 GGAATCTCTGGGCAATTTTATAGATATGATCTCCACTTAATCAATCTCTCATCT 676
QY 641 GCTATATGAATCACTGGAGCTGTGCAGCAGCTCGGAGGACCGACGACGACCA 700
DB 677 GCTATTTCTAGCTGTGAACTCTTTCAGATCTGCCAAACAAACCACTCATGTGA 736
QY 701 GGAAGGACGCACTGGCGAGGCTGCTGCTTACGCTGTGATGCTGTGGCGGTGGTGG 760
DB 737 AATCTGGGTAAACAAAGGCTCTCAACACAAATTTCTTATTATTGTTGTTGTTGTC 796
QY 761 TCTGTTTCAGCCCTTACCACTCAACATCAAGAGTTTATGCGGAGGAGATGCTCCACC 820

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/352,678
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gates, Edward R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: 50801/7044
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1643 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 34..1116
US-09-536-954-3

Query Match 16.7%; Score 184.8; DB 4; Length 1643;
Best Local Similarity 50.6%; Pred. No. 4e-36;
Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;

QY 101 CTCCTGTTCTACACAGCCCTCTGCTCTTACAGTGGGAAACATCTTGGCCCTTTGCC 856
DB 821 TGCCATCC-----TGTCGCGAGCGAGGGCTTTCTTACTGTCCCTTCAGGCCACCGTGG 874
DB 857 CTAATTTCTCGGAATGTAGCCAAAGACATTCGTTCCAGATTCTCTGCACATTTACAGTAT 916
QY 875 CCTCATGAACATGAACATGTGGCATACCCCAATCATTTACTTCTTTCATCCACCCCAAT 934
DB 917 GCCTGATGAACATTCATTTCTGCTGATGGACCCCTTTTATCTACTCTTTTGCATGTAAGSGT 976
QY 935 ACAGGAAATGGCTCTTGGCCATTTTAAAGCTCAAAAGGCTTCTTCTCTCTCTCTCTCT 994
DB 977 ATAAGAGAAAGGTTATGAGGATGCTGAAACGCGCAAGTCAGTGTATCGATTCTTAGTGCTG 1036
QY 995 CCTCTCTCCACCCAGGAAAGCTTCTTCAGAAACACCA 1032
DB 1037 TGAAGTCAGCCCTCGAAGAAATTCACGTGAATGACA 1074

RESULT 5
US-09-016-434-1279
; Sequence 1279, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 5.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREMITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1279:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1643 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: G292056
US-09-016-434-1279

Query Match 16.7%; Score 184.8; DB 4; Length 1643;
Best Local Similarity 50.6%; Pred. No. 4e-36;
Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;

QY 101 CTCCTGTTCTACACAGCCCTCTGCTCTTACAGTGGGAAACATCTTGGCCCTTTGCC 160

```

```
Db 137 CTCGCTATACAGCTGCTCTTCAATCTGGCTCGTGGAACTTACTAGCTTGGTGC 196
Qy 161 TTACTGTCAAAAGAGAGAGATCAACTGCAAGGATCTACTGGTGCACCTGGCTG 220
Db 197 TCATTGTTCAAAAGAGAGAGATCAACTTACACCTCTATTCAACAAATTTGGTGA 256
Qy 221 TGTCTGACCTGCTGTACCTGCGCTTACCGGAAGGTTGGTGTATTGTTGCTGGCT 280
Db 257 TTTCTGATATATTTTACACGCTTTGCTACAGATAGCTACTATGCTATGCTATG 316
Qy 281 CCAGCTGCGCTTTCCGCAAGGGCTTGCAGGCTGACGGGTTTGTCTCTACACGACA 340
Db 317 TTGACTGAGAAATCGAGATGCTTGTGTAGATAACTGCGCTAGTGTATTACATCA 376
Qy 341 CCTAGGGGGGCTTACTCTGATGCTGTGAGCGTGGACATACCAGCTGTGTCT 400
Db 377 CATATGCAAGTGTGAATTTATGACCTTCCCTGAGTATGACCGTTATTGCTGTG 436
Qy 401 GTGCCACTGGGGCGCGCTCCGACGCTGCGCGCGCGAGGCTGTCTGCGTGCGCA 460
Db 437 ACCCTCTACGTACAAAGATAAAAGATTGAACATGCAAAAGGCGTGTGCATATTG 496
Qy 461 TCTGACCTTGGTGTCTGCTGAGAGATGCTTGTCTTGTGATGCCATGACCAAGCGC 520
Db 497 TCTGATGCTAGTATTTGCTCAGACCTCCCACTCTCATCAACCTCTATGTCAAAGCAGG 556
Qy 521 TGTGGGCAAGCTGCTGCTGAGTACAGAGATCAACTGCAAGGATCTACTGCTGGCTG 580
Db 557 AGGCTGAAAGATTGCTGAGTATGCTGAGTATGCTGAGTATGCTGAGTATGCTGAGT 616
Qy 581 TCANGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 640
Db 617 GGAATCTGCTTGGGCGATGTTTATGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 676
Qy 641 GCTATATGAAGATCACTGGAAGCTGTGAGAGATGCTGAGAGATGCTGAGAGATGCTGAG 700
Db 677 GCTATTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 736
Qy 701 GGAAGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 760
Db 737 AATCTGGTGTAAACAAAGGCTCTCAACACAAATTTCTTATTTATTTGTTGTTGTTG 796
Qy 761 TGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 820
Db 797 TCTGTTTCACTTACCACTTACCACTTACCACTTACCACTTACCACTTACCACTTACCACT 856
Qy 821 TGCCATCC-----TGTGCCAGCGAGGGCTTCTTACTGTCTTCTGCTGCTGCTGCTGCTGCTG 874
Db 857 CTAAATTTCTGGAATGTAGCCHAAAGACATTCGTTCCAGATTTCTCTGCACTTTACAGTAT 916
Qy 875 CCTCATGAACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 934
Db 917 GCTGATGAATCTCAATTTGCTGATGAGACCTTTTATCTACTTCTTGTGATGTAAGGGT 976
Qy 935 ACAGGAAATGGCTCTGCTGCTGCTTAAAGCTCAAAAGGCTTCTTCTTCTTCTTCTTCTTCT 994
Db 977 ATAGAGAAAGTTATGAGATGCTCAACGCGAAGTCAAGTATGATGATGATGATGATGATG 1036
Qy 995 CTCTCCACCCAGAGAAAGTTCTTTCAGAAACCA 1032
Db 1037 TGAAGTCAGCCCTGGAAGAAATTCACGTGAATGACA 1074
```

RESULT 6

```
PCT-US93-09636-3
; Sequence 3, Application PC/US9309636
; GENERAL INFORMATION:
; APPLICANT: Birkenbach, Mark
; APPLICANT: Kieff, Elliot
; TITLE OF INVENTION: Epstein Barr Virus Induced Genes
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSER: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/09636
; FILING DATE: Herewith
```

CLASSIFICATION:

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1643 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: linear

FEATURE:

NAME/KEY: CDS

LOCATION: 34..1116

PCT-US93-09636-3

Query Match 16.7%; Score 184.8; DB 5; Length 1643;

Best Local Similarity 50.6%; Pred. No. 4e-36;

Mismatches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;

```
Qy 101 CTCGTCTTACAGAGCCCTCCCTGCTTTCAGTGGCCCTGGGAAACATCTTGGCTTGGC 160
Db 137 CTCGTCTTACAGAGCCCTCCCTGCTTTCAGTGGCCCTGGGAAACATCTTGGCTTGGC 196
Qy 161 TTACTGTCAAAAGAGAGAGATCAACTGCAAGGATCTACTGCTGGCTGCACTGGCTG 220
Db 197 TCATTGTTCAAAAGAGAGAGATCAACTGCAAGGATCTACTGCTGGCTGCACTGGCTG 256
Qy 221 TGTCTGACCTGCTGTTCACGCTGGCTTACCGGGAAGGGTGGTGTGTTATGCTGGCT 280
Db 257 TTTCTGATATCTTTTATCCACGCTTTCCTACAGGATGCTGCTGAGTATGACCGCTTCA 316
Qy 281 CCAGCTGGCTTTCGCAAGGGCTCTGCAAGGCTGACGGCTTGTGCTTACACCGACA 340
Db 317 TTGACTGGAATCGGAGATGCTTGTGAGTAACTGGCTAGTGTGTTTACATCAACA 376
Qy 341 CCTACGGGGGGTCTACTCTCATGCTGCTGTGAGCGTGAACATACCAGCTGTGGTCT 400
Db 377 CATATGCAAGTGTGAATTTATGACCTGCTGAGTATGACCGCTTCAITGCTGTGGTGC 436
Qy 401 GTGCCACTGGGGCGCGCTCCGACGCTGCGCGCGCGAGGCTGGTCTGCTGGCGCA 460
Db 437 ACCCTCTACCTCAACAAAGATAAAGATTGAACATGCAAAAGGCTGTGCATATTG 496
Qy 461 TCTGGAACCTTGGTGTCTGCTGCAAGATGCTTGTCTTGTATGCCCATGACCAAGCGC 520
Db 497 TCTGGAATCTAGTATTTGCTCAGACACTCCCACTCTCATCAACCTTATGTCAAAGCAG 556
Qy 521 TGTGGGCAAGCTGCTGCTGAGTACAGAGATCAACTGCAAGGATCTACTGCTGGCTGCCC 580
Db 557 AGGCTGAAAGATTACATGCTGAGTATCCAAACTTTGAAGAACTAAATCTCTTCTTCT 616
Qy 581 TCATGCTCTGCTGGCTTTCGCAATTTGCTGCGCGAGTGGGATCATCTGCTGCT 640
Db 617 GGATTTCTGCTGGGCAATGTTTATAGTATGCTACTTCCACTTATATCAATCTCATCT 676
Qy 641 GCTATATGAAGATCACTGGAAGCTGTGAGAGATGCTGAGGAGGAGGAGGAGGAGGAGGAG 700
Db 677 GCTATTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 736
```


NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCES/DOCKET NUMBER: PA-0002 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1482:
SEQUENCE CHARACTERISTICS:
LENGTH: 2025 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: 9984506
US-09-016-434-1482

Query Match	10.7%	Score 118.6;	DB 4;	Length 2025;
Best Local Similarity	48.1%	Pred. No. 8.5e-20;		
Matches 399;	Conservative	0;	Mismatches 424;	Indels 6;
				Gaps 2;

Qy	95	TGGCTCTCTGTTCTTACAGACCCCTCTTGCTCTTCACTGCTTGGAGCATCTCTTGCC	154
Db	349	TGCTGCTGTGTCTTACCGCTGGTGTGGGTGTGTCTGAGAGCGCGTGGCGC	408
Qy	155	TTTGGCTTACTCTCAAAGAGCAGGAAGATCAACTGCAAGCATCTACCTGGTGCAAC	214
Db	409	TCTATACATCTTCTGTGCCGCTCAAGACCTGGAATGCTCCACCACATATATGTTCCACC	468
Qy	215	TGGCTGTGTGACCTGCTGTTACCGTGGCCCTTACCGGGAAGGTGGTGTATGTGTC	274
Db	469	TGGCTGTGTGATGCACTGTATCGCGCTCTCCCTGCCGCTGTGGTCTATTACTACGCC	528
Qy	275	TGGCTCCAGTGGCCCTTCGGCAAGGGGCTCTCCAGGCTGACGGGTTTGTGCTCTACA	334
Db	529	CGCGCAGCACTGGCCCTTCAGCACGGTCTCTGCAAGCTGGTGGCTTCTCTTCTACA	588
Qy	335	CCGACCACTACGGGGGGTCTACCTCATGGCTGTGTGACGTGGACCATTACCCAGCTG	394
Db	589	CCAACTTTTACTGAGCATCTCTTCTCACTACCTGCATCAGCGTGCACCGGTGTCTGGCG	648
Qy	395	TGGTCTGTGCCCACTGGGGGCGCGCTCCCGCAGCGGTGGCGCGCAGCGTGGTCTGGC	454
Db	649	TCTTAAGACCTCTGCGCTCCTGCGCTGGGGCGGGCCCGCTACCGTCCGCGGTGGCG	708
Qy	455	TGGCCATCTGACCTTTGGTGTCTGCAGAGCATGCCCTTGCCTTGTATGCCCATGACCA	514
Db	709	GGGCGGTGTGGGTGTGGTGTGCGCTGCCAGGCCCCCGTGTCTACTTTGTGCACACCA	768
Qy	515	AGCGCTGGTGGGAAGCTGGCCCTGCATGAGTAGTACAGCAGCATGGAGTCACTCTCGGC	574
Db	769	CGCGCGCGGGGCGCGGTAACTGCCACGACACCTCGGCACCCGAGCTCTTCAGCGCT	828
Qy	575	TGCCCCCATGGTCTGTGTGGCTTTGCCATTTGCTTGTGGCCAGTGGGATCATCC	634
Db	829	TGTTGGCCCTACAGCTCACTCATGTGGCCCTGCTCTTTCGGGGTGGCTTTGGCGTCA	888
Qy	635	TGTCCTGCTATATGAAGATCACTTGAAGCTGTGCAGCACAGCTCGGGAGGACCCAGTGA	694
Db	889	TTGTCTGTATACGTGTCTATGGCTCGCGACTGCTTAAGCCAGCTACGGGACCTCGGGC	948
Qy	695	CCAGCAGGAAGGACGCCCATGCGGAGGCTGCCTGCTTACCGTCTGATGCTGTGTGGCG	754
Db	949	GGCTCCCTAGGCGCAAGCGCAAGTCCGTGCGCACCATGCG---CGTGGTGTGGCTGTCT	1005
Qy	755	TGGTGGTCTGCTTTACGCCCTTACCACCTCAACATCAAGCAGTTTATGGCGAGAGGATGC	814
Db	1006	TGCGCCTCTGCTTCTCGCCATTCACGTCACCGCACCTCTACTACTCTTCGCTCGC	1065
Qy	815	TTCACTGCGCATCTCTGTGCGAGCGGAGGGCTTTCTTACTGTCCCTTCAGGCCACCGTG	874
Db	1066	TGGACCT---CAGCTGCCACACCTTCAACGGCCATCAACATGGCCTTCAAGGTATACCGCG	1122

Qy 875 CCCTCATGAACATGAACGTGTGGCATTACCCCAATCATTTACTTCTTTGC 923

Db 1123 CGCTGGCCAGTGTCTAACAGTTGGCTTGACCCCGTGTCTACTTCTCTGGC 1171

RESULT 9

```

US-09-170-496D-31
; Sequence 31: Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chien, Derek T.
; APPLICANT: Liaw, Chen W. T.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
; LENGTH: 1020
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-170-496D-31

```

Query Match	10.1%	Score 112;	DB 4;	Length 1020;
Best Local Similarity	48.9%	Pred. No. 2.8e-18;		
Matches	184.	Conservative	0;	Mismatches 375; Indels 27; Gaps 2;

137	QY	TGGAAATACCTTGGCCCTTGGCTTACCTGTTCAAAAGACAGGAAGATCAACTGCACAG	199
140	DB	TTGGCAATACCCCTGGCTCTCTGGGCTTTTCATCGAGACCACAAGTCGGGACCCCGGCCA	199
197	QY	GCATCTACCTGGTGACACCTGGCTGTGTGTGACCTGTGTTACACCGTGGCCTTACCGGAA	256
200	DB	ACGTGTTCCGTGATGCATCTGGCCGTGGCCAGCTTGTGTGGTGTCTGTCTGTGCCACCC	259
257	QY	GGGTGGTGTGTTATGTGTGGGCTCCAGCTGGCCCTTTCCGCAAGGGGCTCTGCAGGGTGA	316
260	DB	GCTTGGTCTACCATTTCTCTGGAAACACATGGCCATTTGGGAAATCGCATGCCGTCTCA	319
317	QY	CGGCGTTTGTGCTCTACACGACACCTACGGGGGGTCTACCTCATGGCCCTGTGTGAGCG	376
320	DB	CCGGCTTCCCTCTTCTACCTCAACATGTACGCCAGCATCTACTCTTCACTTGCATCAGCG	379
377	QY	TGGACCAATTAACCAAGCTGTGCTGTGTGCCCACTGGGGCCCGCCCTCGCGACGGCTGGCC	436
380	DB	CCGACCGTTTCCTGGCCATTGTGCACCCGGTCAAGTCCTCAAGCTCGCAGGCGCCCTCT	439
437	QY	GGCCAGCGCTGTGTGCTGGCCATCTGGACCTTGGTGTCTGTCTGCAGACGATGCCCTTGC	496
440	DB	ACGCAACACTCGCCTGTGCTCTCTGTGGGTGTGTGTGGCTGTGTGCCATGAGCCCGCTGC	499
497	QY	TCTTGATGCCCATGACCAAGCCGTGTGTGGCAAGCTTGGCTGTGATGGAGTACAGACGA	556
500	DB	TGGTGAGCCCAAGACCCCTGCAGACCAACCAACACGCTGGTCTGTCTGCAGCTGTACCGGG	559
557	QY	TGGAGTCACTCCTCGGGCTGCCCTTCATGTGCTCTGTGGCCCTTTTGCCATTGGCTTCTGTG	616
560	DB	AGAAAGC-----CTCCACCATGCCCTGGTGTCTCTGGCAGTGGCCTTCACT	607
617	QY	GGCCAGTGGGGATCATCTCTGTCTGCTATATGAAGATCACTTGGAAAGCTGTGTGCAGCAG	676
608	DB	TCCCGTTTCATACCAACGCTCACTCTCACTGCTGATCATCGCAGCTTCGGCAGGCGCC	667
677	QY	CTCGGGAGACCCAGTGACCAAGGAGGACGCACTGGCGAGGCTGCCTGCTTACGC	736
668	DB	TGCGTGTGGAAGCGCCTCAGACCAAGGCAGTGGCATGCC-----	714
737	QY	TGCTGATGCTGTGTGGCCGTGTGTGTCTGCTTCAAGCCCTTACACCTCAACATCAGCAGT	796
715	DB	--ATAGTGTGGCCATCTTCTGGTCTGCTTCGTGCCCTACACGCTCAACGCTCCGCT	772

Matches 384; Conservative 0; Mismatches 375; Indels 27; Gaps 2;
QY 137 TGGAAACATCTTGGCCCTTTGCCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAG 196
Db 839 TTGGCAATACCTGGCTCTGTGGCTTTTCATCCGAGACCAAGTCGGGACCCCGGCCA 898
QY 197 GCATCTACTGGTGCACCTGCTGTGTGACCTGTCTTACCGTGGCTTACCGGAA 256
Db 899 ACGTGTCTCTGATGATCTGGCCGTTGGCGAGCTTGTGTGGTGTCTGTCTGCCACCC 958
QY 257 GGGTGGTGTGTATGTGTGGCTCCAGCTGGCCCTTTTCGGAAGGGCTCTGCAGGTGA 316
Db 959 GCTGTCTACACCTTCTTGGACCACTGGCCATTTGGGAAATCGCATGCGTCTCA 1018
QY 317 CGGCTTTGTGCTTACACGACCTAGCGGGGTTCTACTCATGGCTGTGTGCGG 376
Db 1019 CGGCTTCTCTTACCTCAACATGTAGCGCAGCATCTACTCTCACCTGCATCAGCG 1078
QY 377 TGGACCATTACCCACTGTGTGTCTGTGCCACTGGGCGCCGCTCGCACGGCTGCCC 436
Db 1079 CGACCGTTTCTGGCCATTTGACCCGGTCAAGTCCCTCAAGCTCGCAGGCCCTCT 1138
QY 437 GCGCAGGCTGTGTGGTGGCCATCTGGACCTTGGTGTCTGTCTGAGAGAGTGCCTTGC 496
Db 1139 AGCCACACTGGCTGTGTCTTCTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1198
QY 497 TCTTATGCCCATGACCAAGCGCTGGTGGCAAGCTGGCTGGCTGATGAGTACAGCAGCA 556
Db 1199 TGGTGAAGCCACAGACCGTGGAGACCAACACAGGTGTGTGTGTGTGTGTGTGTGTGT 1258
QY 557 TGGAGTCACTCTCGGGCTGGCCCTCATGTGTCTGTGTGTGTGTGTGTGTGTGTGTGT 616
Db 1259 AGAAGGC-----CTCCACCATGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1306
QY 617 GGCAGTGGGATCATCTGTCTGTCTGTATGAAGTCACTGGAAGCTGTGAGCAGCAG 676
Db 1307 TCCGTTATCATCAACCGTCACTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1366
QY 677 CTCGGGAGACCCAGTACAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 736
Db 1367 TCGGTGTGGAGAGCGCTCAAGACCAAGGAGTGGCATGTGCG----- 1413
QY 737 TCGTATGTGTGGCGGT 796
Db 1414 --ATAGTGTGGCCATCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1471
QY 797 TCATGCGAGAGGATGTCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 856
Db 1472 AGTGTGCTACTACCGCAGCCTATGGGCTCTGTGCGCACCGCATCTTGGCCCTGG 1531
QY 857 CCTTTCAGGCCACCGTGGCCCTCATGAAATGAACTGTGGCATTACCCCAATCATTTACT 916
Db 1532 CAAACCGCATCACCTCTGCTTCCCTCAGCAGGCTCAAGGGGAGTCTGACCCCATCATGTATT 1591
QY 917 TCTTGG 922
Db 1592 TCTTCG 1597

RESULT 12
US-08-153-848-43
Sequence 43, Application US/08153848
Patent No. 5759804
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: No. 5759804el Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Bicknell
STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER: IBM PC compatible
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/153,848
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 5759804and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31794
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 1901 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 701..1717
US-08-153-848-43
Query Match 10.1%; Score 112; DB 1; Length 1901;
Best Local Similarity 48.9%; Pred. No. 3.5e-18;
Matches 384; Conservative 0; Mismatches 375; Indels 27; Gaps 2;
QY 137 TGGAAACATCTTGGCCCTTTGCCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAG 196
Db 840 TTGGCAATACCTGGCTCTGTGGCTTTTCATCCGAGACCAAGTCCGGGACCCCGGCCA 899
QY 197 GCATCTACTGGTGCACCTGCTGTGTGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 256
Db 900 ACGTGTCTCTGATGATCTGGCGTGGCGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 959
QY 257 GGGT 316
Db 960 GCTGTGTACCACTTCTTGGGAACCACTTGGGCCATTTGGGGAATCGCATGCGTCTCA 1019
QY 317 GCGGTTTGTGCTTACACGACACCTAGCGGGGGTCTTACCTCATGGCTGTGTGTGTGTGT 376
Db 1020 CCGGCTTCTCTTCTTACCTCAACATGTAGCCAGCATCTACTTCTTCCACCTGTACGCG 1079
QY 377 TGGACCATTACCCAGCTGT 436
Db 1080 CGACCGTTTCTTGGCCATTTGTGACCCCGTCAAGTCCCTCAAGCTCCGAGGCCCTCT 1139
QY 437 GCGCAGGCTGT 496
Db 1140 ACGCACACCTGGCTGT 1199
QY 497 TCTTGTATGCCCATGACCAAGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 556
Db 1200 TGGTGAAGCCACAGACCGT 1259
QY 557 TGGAGTCACTCTTGGGCTGCCCTCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 616
Db 1260 AGAAGGC-----CTCCACCATGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1307
QY 617 GCGCAGTGGGGATCATCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 676

1140 ACGCACCTGGCTGTGCTTCTCTGGTGGTGGTGGCTGTGGCCATGGCCCCGGTGC 1199
497 TCTTGATGCCATGACCAAGCCGCTGGTGGGCAAGCTGGCTGCAATGAGTACAGCAGCA 556
1200 TGGTGAACCCACAGACCGTGCAGACCAACACACGGTGGTCTGGCTGCAGCTGTACCGGG 1259
557 TGGAGTCAGTCTCTGGGCTGCCCTCATGGTCTCTGGTGGCTTTTGCCATTGGCTTCTGTG 616
1260 AAGAGC-----CTCCACCATGGCTGGTGTCCCTGGCAGTGGGCTTCACT 1307
617 GGCAGTGGGGATCATCTGTCTCTGTATATGAAGTACACTGGAGCTGTGCAGCAG 676
1308 TCCGTTTCATCACACAGGTCACTGTCTACTCTGTATCATCCGAGCTTGGCGAGGGCC 1367
677 CTCGGAGGACCCAGTGACAGCAGGAAGAGCGCACCTGGCGAGGCTGCCTGTCTTACGC 736
1368 TCGCTGTGGAGAGCGCTCAAGACCAAGCAGTGGCATGCGC----- 1414
737 TCGTGATCTGGTGGCGGTGGTCTGTCTGTCTAGCCCTTACACCTCAACATCAAGCAGT 796
1415 --ATAGTCTGGCCATCTTCTCTGGTCTGTCTGTCTGTGGCTTACCAGCTCAACCGCTCG 1472
797 TCATGGCGAGAGGATGCTCCACTGCGCATCTGTGCGGAGCGAGGGCTTCTTACTGT 856
1473 AGTGTCTCACTACCGAGCGCATGGGGCTCTCTGCGCCACCCAGCGCATCTTGGCCCTGG 1533
857 CCGTTTCAGGCCACCGTGGCCCTCATGAACATGAATGTGGCATATCCCAATCATTTACT 916
1533 CAACCGCATCACTCTCTGCTCTGCTCTGCTTCCAGCGCTCAACGGGCACTCGACCCCATGTATT 1599
917 TCTTTG 922
1593 TCTTCG 1598

RESULT 15
PCT-US93-11153-43
? Sequence 43, Application PC/TUS93111153
? GENERAL INFORMATION:
? APPLICANT: Godiska, Ronald
? APPLICANT: Gray, Patrick W.
? APPLICANT: Schweikart, Vicki L.
? TITLE OF INVENTION: Novel Seven Transmembrane Receptors
? NUMBER OF SEQUENCES: 64
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
? ADDRESSEE: Bicknell
? STREET: 6300 Sears Tower, 233 South Wacker Drive
? CITY: Chicago
? STATE: Illinois
? COUNTRY: USA
? ZIP: 60606
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US93/11153
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/977,452
? FILING DATE: 17-NOV-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Noland, Greta E.
? REGISTRATION NUMBER: 35,302
? REFERENCE/DOCKET NUMBER: 31794
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (312) 474-6300
? TELEFAX: (312) 474-0448
? TELEX: 25-3856
?

; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1901 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 701..1717
PCT-US93-11153-43

Query Match 10.1%; Score 112; DB 5; Length 1901;
Best Local Similarity 48.9%; Pred. No. 3.5e-18;
Matches 384; Conservative 0; Mismatches 375; Indels 27; Gaps 2;
QY 137 TGGGAAACATCCTTGGCCCTTGGCTTACCTGTCAAAAGAGCAGGAGATCAACTGCACAG 196
Db 840 TTGGCAATACCTTGGCTCTGTGGCTTTTCATCGAGACCAAGTCCGGGACCCCGGCCA 899
QY 197 GCATCTACCTGGTGCACCTGGCTGTCTGTGACCTGTCTTACCGTGGCTTACCGGAA 256
Db 900 ACGTGTTCCTGTGATCTGTGGCCGTGGCCGACTTGTGCTGCTGTGCTGTGCTGCTG 959
QY 257 GGGTGGTGTGTATGTCTGGGCTCCAGCTGGGCTTTCCGCAAGGGGCTCTGCAGGCTGA 316
Db 960 GCCTGGTCTACCACTTCTCTGGGAACCACTTGGGCAATTTGGGAAATCGCATGCCGTCTCA 1019
QY 317 CGGCGTTTGTCTTACACCGACACCTACGGGGGGGTCTACCTCATGGCTGTGTGACGG 376
Db 1020 CCGGCTTCCCTTCTTACTCAACATGTACGCCAGCATCTACTTCTCTACCTGCTCATCAGCG 1079
QY 377 TGGACCAATTACCCAGCTGTGTCTGTGCCACTTGGGGCCCGGGCTTCCGCAAGGGTGGCC 436
Db 1080 CCGACCGTTTCTGGCCATTTGTGCAACCGGTCAAGTCCCTCAAGCTCCGCAAGCCCTCT 1139
QY 437 GCGCAGCTGTGTCTGGCGGCATCTGGACCTTGTGTGTGTGTGTGTGTGTGTGTGTGTG 496
Db 1140 ACGCACACCTTGGCTGTGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1199
QY 497 TCTTGATGCCCATGACCAAGCGGTGTGTGGGCAAGCTGGGCTGCTGATGGAGTACAGCAGCA 556
Db 1200 TGGTGAAGCCACAGACCGTGCAGACCAACACACGCTGTGTGTGTGTGTGTGTGTGTGTGTG 1259
QY 557 TGGAGTCAGTCTCGGGCTGCCCTCATGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 616
Db 1260 AGAAGGC-----CTCCACCATGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1307
QY 617 GGCAGTGGGGATCATCTGTCTGTCTATATGAAGATCACCTGGAAGCTGTGCAGCAGAG 676
Db 1308 TCCGTTTCATCACACCGTCACTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1367
QY 677 CTCGGAGAGCCAGTACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 736
Db 1368 TGCCTGTGGAGAGCGGCTCAAGACCAAGGAGTGTGGCATGATCGCC----- 1414
QY 737 TGCTGATGTGGTGGCGGCTG 796
Db 1415 --ATAGTGTGGCCATCTTCTG 1472
QY 797 TCATGGGAGAGGAGTG 856
Db 1473 ACGTG 1532
QY 857 CCGTTTACGGCCACCGTGGCCCTCATGAACATGAACATGAACATGAACATGAACATGAACATGAAC 916
Db 1533 CAACCGCATCATCTCTCTG 1592
QY 917 TCTTTG 922
Db 1593 TCTTCG 1598

Search completed: February 21, 2004, 21:28:25
Job time : 115 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 19:18:54 ; Search time 467 Seconds

(without alignments)
8277.977 Million cell updates/sec

Title: US-10-041-615-33

Perfect score: 1104

Sequence: 1 atgataaacaggccctta.....gtgtcatttgctttatgact 1104

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2308684 seqs, 175082206 residues

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: Published Applications NA:*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/2/pubpna/ECT_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/2/pubpna/ECTUS_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1104	100.0	1104	15	US-10-041-615-33
2	779	70.6	4319	14	Sequence 33, Appl
3	779	70.6	4319	15	Sequence 1619, Ap
4	672.8	60.9	1526	9	US-10-292-798-1293
5	397.2	36.0	536	10	Sequence 40, Appl
6	378.4	34.3	556	11	Sequence 35, Appl
7	184.8	16.7	1086	14	Sequence 64, Appl
8	184.8	16.7	1086	15	Sequence 1, Appl
9	184.8	16.7	1643	15	Sequence 77, Appl
10	184.8	16.7	2477	14	Sequence 1279, Ap
11	183.2	16.6	1086	14	Sequence 39, Appl
12	183.2	16.6	1638	14	Sequence 205, App
13	172.2	15.6	1556	15	Sequence 111, App
14	148	13.4	594	9	Sequence 29, Appl
15	123.4	11.2	548	10	Sequence 90, Appl
					Sequence 10498, A

16	119.4	10.8	1020	14	US-10-023-586B-3
17	119.4	10.8	1076	16	US-10-333-946-20
18	118.6	10.7	2025	14	US-10-225-567A-216
19	118.6	10.7	2025	14	US-10-101-510-722
20	118.6	10.7	2025	15	US-10-305-720-1482
21	117	10.6	2118	14	US-10-101-510-431
22	115	10.4	1017	10	US-09-885-453-3
23	115	10.4	1017	14	US-10-079-384-5
24	112	10.1	1020	9	US-09-788-133-1
25	112	10.1	1020	14	US-10-251-385-31
26	112	10.1	1020	14	US-10-251-385-181
27	112	10.1	1900	15	US-10-305-720-1484
28	112	10.1	2070	14	US-10-225-567A-303
29	111.8	10.1	1082	14	US-10-023-586B-1
30	111.8	10.1	1334	14	US-10-017-161-707
31	111.8	10.1	1334	15	US-10-292-798-619
32	111.8	10.1	3143	14	US-10-275-910-1
33	110.2	10.0	1002	14	US-10-088-726-25
34	106.8	9.7	2218	9	US-09-214-904-3
35	106.8	9.7	2219	14	US-10-112-599A-1
36	106	9.6	3733	14	US-10-005-956-844
37	105.4	9.5	1176	11	US-09-826-509-462
38	105.4	9.5	1378	15	US-10-305-720-1123
39	105.4	9.5	3733	14	US-10-005-956-11
40	105.4	9.5	3733	14	US-10-005-956-563
41	105.4	9.5	3733	14	US-10-005-956-565
42	105.4	9.5	3733	14	US-10-005-956-567
43	105.4	9.5	3733	14	US-10-005-956-850
44	105.4	9.5	3733	14	US-10-225-567A-47
45	105.2	9.5	1829	9	US-09-823-114-7

ALIGNMENTS

RESULT 1

US-10-041-615-33
; Sequence 33, Application US/10041615
; Publication No. US20040014038A1
; GENERAL INFORMATION: Stacie J
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Ellerman, Karen
; APPLICANT: Smithson, Glennda
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Padigaru, Muraliidhara
; TITLE OF INVENTION: No. US20040014038A1el GPCR-Like Proteins and Nucleic Acids Encodi
; FILE REFERENCE: 21402-233-061
; CURRENT APPLICATION NUMBER: US/10/041.615
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: 60/259,552
; PRIOR FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: 60/260,544
; PRIOR FILING DATE: 2001-01-09,277,405
; PRIOR APPLICATION NUMBER: 60/277,405
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: Curaseq version 0.1
; SEQ ID NO 33
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1080)
US-10-041-615-33

Query Match 100.0%; Score 1104; DB 15; Length 1104;
Best Local Similarity 100.0%; Pred. No. 2.6e-310;
Matches 1104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGATAAACTAGGCCCTTACTTCACACCCACCAAAATAAACTAAGATTAAAGAT 60
|||||

1	ATGATAAAC	TAGGCCCTT	ACTTCACAC	CCACAAA	ATAA	AAAACT	TAA	AGAT	60
61	TTAAATG	GGAAAG	CAAACT	TGTGAA	ATTTTT	TGGCT	CTCTG	TCTTAC	120
61	TTAAATG	GGAAAG	CAAACT	TGTGAA	ATTTTT	TGGCT	CTCTG	TCTTAC	120
121	CTGGTCTT	CAGTGC	CTGGGAAA	CATCCTT	GGCCCTT	TGCTT	ACCTG	TCAAAAG	180
121	CTGGTCTT	CAGTGC	CTGGGAAA	CATCCTT	GGCCCTT	TGCTT	ACCTG	TCAAAAG	180
181	AGATCAAT	GCACAGG	CACTA	CTGGT	GGCACTGG	CTGTG	CTGAC	CTGCTGTT	240
181	AGATCAAT	GCACAGG	CACTA	CTGGT	GGCACTGG	CTGTG	CTGAC	CTGCTGTT	240
241	GTGGCTT	TACCGG	GAAGG	TGGTGT	TATGTG	TGTGG	GTTC	CAGCTGG	300
241	GTGGCTT	TACCGG	GAAGG	TGGTGT	TATGTG	TGTGG	GTTC	CAGCTGG	300
301	GGGCTCTG	CAGGCT	GACGG	CTTTG	TGCTCTA	CACG	CACTA	CGGGGG	360
301	GGGCTCTG	CAGGCT	GACGG	CTTTG	TGCTCTA	CACG	CACTA	CGGGGG	360
361	ATGGCTGT	GTGAG	CGTG	GCACAT	TACCA	CGTGT	GTGTG	CGCCAC	420
361	ATGGCTGT	GTGAG	CGTG	GCACAT	TACCA	CGTGT	GTGTG	CGCCAC	420
421	CTCCGAC	CGGCTGG	CGCG	CCAG	CGCTGG	CTGG	CCATCTG	AGACCT	480
421	CTCCGAC	CGGCTGG	CGCG	CCAG	CGCTGG	CTGG	CCATCTG	AGACCT	480
481	CAGCGATG	CCCTT	CTCTT	GATGCC	ATGACCA	ACGCG	CTGGT	GGGCAAG	540
481	CAGCGATG	CCCTT	CTCTT	GATGCC	ATGACCA	ACGCG	CTGGT	GGGCAAG	540
541	ATGGAGT	CAGCAG	CATG	AGTCA	GCTCT	CGGG	CTGCC	CTCAT	600
541	ATGGAGT	CAGCAG	CATG	AGTCA	GCTCT	CGGG	CTGCC	CTCAT	600
601	GCCATTG	GCCTT	CTGTG	GCC	CAGTGG	GATCAT	CTGT	CTCTAT	660
601	GCCATTG	GCCTT	CTGTG	GCC	CAGTGG	GATCAT	CTGT	CTCTAT	660
661	AGACTGT	CAGCAG	CACTG	CGG	AGAC	CCAGT	GCAG	GAAG	720
661	AGACTGT	CAGCAG	CACTG	CGG	AGAC	CCAGT	GCAG	GAAG	720
721	GGCTGC	CTTAC	GTCTG	ATCTG	TGG	CCGTGG	TGTGT	CTGCTT	780
721	GGCTGC	CTTAC	GTCTG	ATCTG	TGG	CCGTGG	TGTGT	CTGCTT	780
781	CTCAAC	ATCAAG	CACTT	CATG	CGCAG	AGGAT	GTCTC	ACTCC	840
781	CTCAAC	ATCAAG	CACTT	CATG	CGCAG	AGGAT	GTCTC	ACTCC	840
841	AGGGCTTT	CTACT	GTCC	CTT	CAGG	CCACCG	CTCAT	GAACAT	900
841	AGGGCTTT	CTACT	GTCC	CTT	CAGG	CCACCG	CTCAT	GAACAT	900
901	ACCCAA	TCATTT	ACTTCT	TTTG	CA	TCCAC	CCAT	TACGGA	960
901	ACCCAA	TCATTT	ACTTCT	TTTG	CA	TCCAC	CCAT	TACGGA	960
961	AGCTCAA	AGG	CTTCT	CTCT	CTCT	CTCT	CTCT	CTCA	1020
961	AGCTCAA	AGG	CTTCT	CTCT	CTCT	CTCT	CTCT	CTCA	1020
1021	TCAAGAA	CA	CCAA	GTAT	CAC	CCG	CCAG	AGT	1080
1021	TCAAGAA	CA	CCAA	GTAT	CAC	CCG	CCAG	AGT	1080
1081	TAA	CGT	GT	CA	TTT	GT	AT	GA	1104
1081	TAA	CGT	GT	CA	TTT	GT	AT	GA	1104

RESULT 2

```

US-10-017-161-1619
; Sequence 1619, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1619
; LENGTH: 4319
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(4319)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(308)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3128)..(3531)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3660)..(4119)
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (1)..(48)
; OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-1619

```

Query Match	70.6%	Score 779;	DB 14;	Length 4319;
Best Local Similarity	83.6%;	Prod. No. 1.6e-215;		
Matches 1006;	Conservative 0;	Mismatches 0;	Indels 197;	Gaps 3;
QY	97	CGCTCTCTGTTCTACACAGCCCTCTGCTTTCAGTGCCTCGGAAACATCTCTTGCCCTT	156	
Db	2940	CGCTCTCTGTTCTACACAGCCCTCTGCTTTCAGTGCCTCGGAAACATCTCTTGCCCTT	2999	
QY	157	TGCCTTACCTCTCAAAAGACGGAAGATCAACTGCACAGGCATCTACTCGTGACCTG	216	
Db	3000	TGCCTTACCTCTCAAAAGACGGAAGATCAACTGCACAGGCATCTACTCGTGACCTG	3059	
QY	217	GCTGTGTCAGCTGCTGTTTCACTGCTGGCTTACCGGGAAGGTCGCTGTATTGCTG	276	
Db	3060	GCTGTGTCAGCTGCTGTTTCACTGCTGGCTTACCGGGAAGGTCGCTGTATTGCTG	3119	
QY	277	GGCTCCAGCTGGCCCTTTTCGGCAAGGGGCTCTGCAGGCTGACGGCTTTGTGCTCTACACC	336	
Db	3120	GGCTCCAGCTGGCCCTTTTCGGCAAGGGGCTCTGCAGGCTGACGGCTTTGTGCTCTACACC	3179	
QY	337	GACACTACGGGGGGGTCTACTCTATGGCCCTGTGTAGCGTGGACCATTACCCAGCTGTG	396	
Db	3180	GACACCTAC - GGGGGGGTCTACTCTATGGCCCTGTGTAGCGTGGACCATTACCCAGCTGTG	3238	
QY	397	GTCTGTGCCACTGGGGCCCGCGCTCCGCAACGGCTGGCGCGCCAGCGCTGCTTGGCGTG	456	
Db	3239	GTCTGTGCCACTGGGGCCCGCGCTCCGCAACGGCTGGCGCGCCAGCGCTGCTTGGCGTG	3298	
QY	457	GCCATCTGGACCTTGTGTGCTGTGCAGACGATGCCCTTGTCTTTGATGCCATGACCAAG	516	
Db	3299	GCCATCTGGACCTTGTGTGCTGTGCAGACGATGCCCTTGTCTTTGATGCCATGACCAAG	3358	

Db 1440 CATCCACCATTACAGGAAATGGCTCTGGSCATTTTAAAGTCAAGGGCTTCTCTCT 1499

Qy 983 CCTCTCTCTCTCTCTCTCTCC 1002

Db 1500 CCTCTCTCTCTCTCTCTCACC 1519

RESULT 5

US-09-791-932-35
; Sequence 35, Application US/09791932
; Publication No. US20030003451A1
; GENERAL INFORMATION:
; APPLICANT: Vogeli, Gabriel
; APPLICANT: Parodi, Luis A.
; APPLICANT: Hiebsch, Ronald R.
; APPLICANT: Lind, Peter
; APPLICANT: Kaytes, Paul S.
; APPLICANT: Ruff, Valerie
; APPLICANT: Ruff, Rita M.
; APPLICANT: Wood, Linda S.
; TITLE OF INVENTION: No. US20030003451a1el G Protein-Coupled Receptors Cross-Referen
; FILE REFERENCE: 00325.US1
; CURRENT APPLICATION NUMBER: US/09/791,932
; CURRENT FILING DATE: 2001-02-23

```

; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,304
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,303
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,397
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,247
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/188,880
; PRIOR FILING DATE: 2000-03-13
; PRIOR APPLICATION NUMBER: 60/217,369
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/217,370
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/218,492
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 60/186,810
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: 60/188,064
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: 60/186,457
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: 60/213,861
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/194,344
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/218,337
; PRIOR FILING DATE: 2000-07-14
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 35
; LENGTH: 536
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-791-932-35

```

Query Match	36.0%;	Score 397.2;	DB 10;	Length 536;
Best Local Similarity	97.5%;	Pred. No. 6e-105;		
Matches 435:	Conservative	0: Mismatches	8: Indels	3: Gaps

Qy	285	CTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTTGTCCTACACCGACACTA	344
Db	8	CTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTTGTCCTACACCGACACTA	67

Db 8 CTGSCCTTTCCGCAAGGGGCTCTCGAGGCTGACGGGTTGTGCTCTACACGACACCTA 67

Ov 345 CGGGGGGGTCTACCTCA TGGCCTGTGTGACGCTGCAACCAATTACCGAGTGTGGTCTGTGC 404

345 CGGGGGGGTCTACCTCATGGCCCTGTGAGCGTGGACCATACCCAGCTGTGGTGTGTC 404

68	C	-GGGGGCTTACCTCATGTGGCCCTGTGTAGCGTGGACCATTA	CCACAGCTGTGCTCTGTGC	126
405	QY	CCACTGGGGCCGCGCCTCCGCA	CGGCTGGCCGCGCAGGCTGGTCTGGCGTGGCCATCTG	464
127	Db	CCACTGGGGCCCGTGCCTCCGCA	CGGCTGGCCGCGCCAGGCTGGTCTGGCTGGCCATCTG	186
465	QY	GACCTTGTGTGCTGTGCAGACGAT	GCCTTGTCTTGATGCCATGACCAAGCGCGCTGGT	524
187	Db	GACCTTGTGTGCTGTGCAGACGAT	GCCTTGTCTTGATGCCATGACCAAGCGCGCTGGT	246
525	QY	GGGCAAGCTGGCCTGCATGGAGTACACAGCATGGATCAGTCTCTC	-GGCGTCGCCCTCA	583
247	Db	GGGCAAGCTGGCCTGCATGGAGTACACAGCATGGAGTCA	GTCTCTCGGGGCTGCCCTCA	306
584	QY	TGCTCCTGTGGCCTTTGCCATTGGCTTCTGTGGCCAGTGGGGA	TCATCTGTCTCTGCT	643
307	Db	TGCTCCTGTGGCCTTTGCCATTGGCTTCTGTGGCCAGTGGGGA	TCATCTGTCTCTGCT	366
644	QY	ATATGAAGTACACTTGGGAAGCTGTGCAGCAGCTCGGGAGGACCCAGT	CACACAGCAGGA	703
367	Db	ATATGAAGTACACTTGGGAAGCTGTGCAGCAGCT	-GGGAGACCCAGTACACAGCGGA	425
704	QY	AAGAGCGCCACTTGGGAGGCTGCCTG	729	
426	Db	AAGGACACACCGCGGGGACGCCG	451	

PRECIT.T 6

```

US-09-801-944B-64/c
; Sequence 64, Application US/09801944B
; Publication No. US20040014169A1
; GENERAL INFORMATION:
; APPLICANT: Vogeli, Gabriel
; APPLICANT: Wood, Linda S.
; TITLE OF INVENTION: No. US20040014169A1e1 G Protein-Coupled Receptors
; FILE REFERENCE: 00100US1
; CURRENT APPLICATION NUMBER: US/09/801,944B
; CURRENT FILING DATE: 2003-08-01
; PRIOR APPLICATION NUMBER: 60/187,828
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,715
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,929
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,930
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,825
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,833
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,830
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,829
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,582
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,581
; PRIOR FILING DATE: 2000-03-08
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 64
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-801-944B-64

```

Query Match 34.3%; Score 378.4; DB 11; Length 556;
Best Local Similarity 98.8%; Pred. No. 1.8e-99;
Matches 403; Conservative 0; Mismatches 1; Indels 4; Gaps 2

	Matches	403;	Conservative	0;	Mismatches	1;	Indels	4;	Gaps	2
y	697	AGCAGGAAAGGACGCCACTGGCGAGGCTGCCTGCTTACGCTGTGATGCTGTGGCCGTG	756							

697 AGCAGGAAAGGACGCCACTGGCAGGCTGCCTGCTGCTGATGCTGTGGCCGCTG 756

Db 556 ACAGGAAAGACGCGACCTGGGAGGCTGCTGCTTACCGTCTGATGCTGGTGGCCGTG 497
Qy 757 GTGGTCTGCTTACGCCCCCTACACCTCAACATCAACAGCAGTTTATGGCGAGAGGATGTC 816
Db 496 GTGGTCTGCTTACGCCCCCTACACCTCAACATCAACAGCAGTTTATGGCGAGAGGATGTC 437
Qy 817 CACCTGCGATCCTGTGCGGAGCGAGGCTTCTTACTGCTCCCTTCAGGCCACCGTGGCC 876
Db 436 CACCTGCGATCCTGTGCGGAGCGAGGCTTCTTACTGCTCCCTTCAGGCCACCGTGGCC 377
Qy 877 CTATGAGACATGACTGTGGCATCCCAATCAATTTACTTCTTGGCATCCACCCATTAC 936
Db 376 CTATGAGACATGACTGTGGCATCCCAATCAATTTACTTCTTGGCATCCACCCATTAC 318
Qy 937 AGAAATGGCTCTGGGCAATTTAAAGCTCAAGGGTCTTCTCTCTCTCTCTCTCTCTCT 996
Db 317 AGAAATGGCTCTGGGCAATTTAAAGCTCAAGGGTCTTCTCTCTCTCTCTCTCTCTCT 261
Qy 997 TCTTCCACCCAGGAAAGCTTCTTCAAGACACAGATATCAACCCGAGGAGGTTCT 1056
Db 260 TCTTCCACCCAGGAAAGCTTCTTCAAGACACAGATATCAACCCGAGGAGGTTCT 201
Qy 1057 ATGTTCTTAGCGAGCAGCTGGTCTAACCGTGTCAATTTGCTTTATGACT 1104
Db 200 ATGTTCTTAGCGAGCAGCTGGTCTAACCGTGTCAATTTGCTTTATGACT 153

RESULT 7

US-10-222-024-1
; Sequence 1, Application US/10222024
; Publication No. US2003010487A1
; GENERAL INFORMATION:
; APPLICANT: Pfizer Ltd. (GB)
; APPLICANT: Pfizer Inc. (EP except GB, US, JP)
; TITLE OF INVENTION: Neuropeptide receptor and uses thereof
; FILE REFERENCE: PCS22032
; CURRENT APPLICATION NUMBER: US/10/222,024
; CURRENT FILING DATE: 2002-08-15
; PRIOR FILING DATE: 2001-08-15
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1086
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-222-024-1

Query Match 16.7%; Score 184.8; DB 14; Length 1086;
Best Local Similarity 50.6%; Pred. No. 4.8e-43;
Matches 475; Conservative 0; Mismatches 457; Indels 5; Gaps 1;
Qy 101 CTCTGTTCTACAGCGCCCTCTGGTCTTCAAGTCTGAGTCCCTGGGAAACATCTTGGCCCTTTGCC 160
Db 104 CTCTGCTATTACGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 163
Qy 161 TTACTCTCAAGAGCAGAGAAATCAACTGACAGGCTATCTACCTGGTGGACCTGGCGTG 220
Db 164 TCATGTTTCAACAGAGAAATCACTTACACCTCTTATCAACAAATTTGGTGA 223
Qy 221 TGCTGACCTGCTTTCACCGTGGCTTACCGGAGAGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 280
Db 224 TTTCTGATATATCTTTTACACGCGCTTTCCTACAGAAATAGCCTACTATGCAATGGGCT 283
Qy 281 CCAGCTGGCTTTTCGGCAAGGGCTCTGCAAGGCTGACGGCTTGTGCTCTACACCGACA 340
Db 284 TTGACTGGAGATCGGAGATGCTTGTGTAGGATTAAGTGGCTAGTGTGTTTACATCAACA 343
Qy 341 CTTACGGGGGGTCTTACTCATGGCTGTGTGAGCGTGGACACATTACCGAGTGTGGTCT 400
Db 344 CATATGAGGTTGTGAATTTATGACCTGCTGAGTATTGACCGCTTCATTGCTGTGGTGC 403
Qy 401 GTGCCACTGGGGCCCGCGCTCGGACAGGCTGGCGCGCGAGGCTGGTGTCTGCTGGGCA 460

Db 404 ACCCTCTAGCTACACAAAGATAAAAAGATTGACATGCAAAAGGCGTGTGCATATTTG 463
Qy 461 TGTGACCTTGT 520
Db 464 TGTGATTTCTAGTATTTGTCTCAGACACTCCCACTCTCTCATCAACCCCTATGTCAAAAGCAGG 523
Qy 521 TGTGTGGCAAGCTGGCTGTGATGGAGTACAGCAGCATGTGAGTCTCTGGGTGCCCC 580
Db 524 AGCTGAAGAGNATACATGATGGAGTATCCAACTTTTGAAGAACTAAATCTCTTCCCT 583
Qy 581 TATGCTCTGT 640
Db 584 GGATTTCTGT 643
Qy 641 GCTATATGAAGATCACTGGAAGCTGTGAGCAGCAAGCTCGGAGGAGACCCAGTGAACAGCA 700
Db 644 GCTATTTCTCAGATCTGCTGCAACTCTTTCAGAACTGCAAAACCAAAACCCACTCACTGAGA 703
Qy 701 GGAAGAGCCACTGTGGAGGCTGT 760
Db 704 AATCTGTGTGAACAAAGAGGCTCTCAACAAATTTATTTATTTATTTATTTATTTATTTATTTATTT 763
Qy 761 TGTGCTTACAGCCCTTACCACTCAACATCAAGCAGTTTCAAGCGAGAGGAGTGTCCACC 820
Db 764 TGTGTTTACACCTTACCATGTTGCAATTTATTCACATATGATTAAGAAGCTTCGTTTCT 823
Qy 821 TGGCATCC-----TGTGCGGAGCGAGGGCTTTTCTTACTGCTTCTTCTTCTTCTTCTTCTTCTTCT 874
Db 824 CTAAATTTCTTGAATGTAGCCAAAGACATTCGTTCCAGATTTCTCTGACATTTTACAGTAT 883
Qy 875 CCCTCATGAACATGAACCTGTGGCATTAACCCCAATCATTTACTTCTTGTTCATCCACCCATT 934
Db 884 GCCTGATGAATTCATTAATGTCTGATGGACCTTTTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 943
Qy 935 ACAGGAAATGGCTCTGGGCAATTTAAAGCTCAAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 994
Db 944 ATAAAGAAAGGTTATGAGATGCTGAAACGCGCAAGTCAAGTATGATGATGATGATGATGATGATGATG 1003
Qy 995 CTCTCTCCACCCAGGAAAGCTTTTTCAGAAACACCA 1032
Db 1004 TGAAGTCAGCCCTGGAAGAAATTCACGTGAATGACA 1041

RESULT 8

US-10-251-385-77
; Sequence 77, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
; TITLE OF INVENTION: Protein-Coupled
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/10/251,385
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US/09/170,496
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 77
; LENGTH: 1086
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-251-385-77

Query Match 16.7%; Score 184.8; DB 14; Length 1086;
Best Local Similarity 50.6%; Pred. No. 4.8e-43;
Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;
Qy 101 CTCTGTTCTACAGCGCCCTCTGGTCTTCAAGTGTGCTTCAAGTGTGCGGAAACATCTTGTCCCTTTGCC 160

104 CTCTGCAATACAGCCCTCGTCTTCACTATGGGCTCGTGGAACTTACTAGCCTTGGTCG 163
161 TTACCTGTCAAAAGAGCAGGAGATCAACTGCACAGGCACTTACCTGTGCACTGGCTG 220
164 TCATTGTTCAAAACAGGAAATAAATCACTTACACCCCTCTATTCAACAATTTGGTGA 223
221 TGTCTGACCTGTCTTACCGTGGCTTACCGGGAAGGGTGTGTTATGTGTGGCT 280
224 TTCTCTGATATCTTTTACACGGCTTTGCCCTTACACGGAATAGCCTTACTATCAATGGGCT 283
281 CCAGCTGGGCTTTTGGCAAGGGCTCTGACGCTCAGCGCTGCTGTCTTACACCGGCA 340
284 TTGACTGGGAATCGAGATGCTTGTGTAGATTAATCGCGTAGTGTATTACATCAACA 343
341 CTTACGGGGGGTCTTACCTATGGCTGTGTGAGCGTGGACCATPACCCAGTGTGTC 400
344 CATATGCAAGTGTGAATTTATGACCTGCTGAGTATTGACCGCTTCAATGCTGTGGTC 403
401 GTGCCACTGGGGCCGCGCTCCGACGGCTGGCGCGCGCAGGCTGCTGCTGGGCA 460
404 ACCCTTACGCTTACAAAGATTAAGAGATTGAACATGCAAAAGCGGTGTGCAATT 463
461 TCTGGACCTTGGTGTCTGACAGGATGCCCTTCTGTATGCCCATGACCAAGCGC 520
464 TCTGGATTCTAGTATTTGTCTGACACACTCCCACTCTCATCAACCTTAATGTC 523
521 TGGTGGCAAGCTGGCTGCTGATGAGTACAGCAGTGGAGTCACTCTCGGGCTGCCCC 580
524 AGGCTGAAAGGATACATGATGAGGATGTCAAACTTTTGAAGAACTAAATCTCTTCCCT 583
581 TCATGCTGCTGGTGGCTTTGCAATGGCTTCTGTGGCGAGTGGGGATCATCTGTCTCT 640
584 GGATCTGCTTGGGCAATGTTTCATAGGATATGTACTTCCACTTATATCATCTCATCT 643
641 GCTATATGAATACCTGAGGAGTGTGACGACAGCTCGGGAGGACCCAGTGACGACA 700
644 GCTATTCTCAGATCTGCTCAAACTCTTCAGAACTGCCAAACAAAACCCACTCACTGAGA 703
701 GGAAGGAGCGCACTGGCGAGGCTGCTGCTTACGCTGTGATGCTGTGTGGCGGTGG 760
704 AATCTGGTGAACAAAGGCTCTCAACACAATTTCTTATTTATTTGTTGTTGTC 763
761 TGTGCTTACGCTTACCACTCAACATCAAGAGTTCATGCGGAGAGGATGCTCCACC 820
764 TCTGTTTACACCTTACCACTTGTGCAATTTATCAACATATGATTAAGAGCTTCTGTTCT 823
821 TGCCATCC-----TGTGCGGAGCGGAGGCTTCTTACTGTCCTTCAGGCCACGTTG 874
824 CTAAATTTCTGGAATGAGCAAGACATTCGTTCCAGATTTCTTGCACTTTACAGTAT 883
875 CCTCATGAACATGAATGCGCATTTACCCCAATCATTTTACTTCTTTGCACTCCACCAT 934
884 GCCTGATGAATTTCAATTTGCTGATGAGCCCTTTTATCTACTTCTTTGCAATGAAGGT 943
935 ACAGGAATGCTCTGGCAATTTTAAAGCTCAAGGCTTCTTCTCTCTCTCTCTCTCT 994
944 ATAAGAGAAGGTTATGAGGATGCTGAAACGGCAAGTCAGTGTATCGATTTCTAGTCTG 1003
995 CCTCTCCACCCGAGGAAAGCTTCTTTCAGAAACACCA 1032
1004 TGAAGTCAGCCCTTGAAGAAATTCACGTGAATGACA 1041

RESULT 9
US-10-305-720-1279
; Sequence 1279, Application US/10305720
; Publication No. US20040010136A1
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice K.; Seilhamer, Jeffrey J.
; FILE OF INVENTION: Composition for the Detection of Signaling Pathway Gene Expression
; FILE REFERENCE: PA-0002-1 CON
; CURRENT APPLICATION NUMBER: US/10/305,720

; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/016,434
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 1490
; SOFTWARE: PERL Program
; SEQ ID NO 1279
; LENGTH: 1643
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: GenBank ID No. US20040010136A1 g292056
US-10-305-720-1279

Query Match 16.7%; Score 184.8; DB 15; Length 1643;
Best Local Similarity 50.8%; Pred. No. 5.8e-43;
Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;
101 CTCTGTTCTACACAGCCCTCCTGGTCTTCAAGTCCCTGGGAAACATCTTGGCCCTTTGCC 160
137 CTCTGCATTACAGCCCTCGTTCATCATTTGGGCTCGTGGAACTTACTAGCCTTGGTCG 196
161 TTACCTGTCAAAAGAGAGGAGATCAATGACAGGCACTTACCTGGTGGCACTGGCTG 220
197 TCATTGTTCAAAACAGGAAAAAATCAACTCTACCAACCTCTATTCAACAATAATTTGGTGA 256
221 TGTCTGACCTGCTGTTTACCGTGGCTTACCGGGAAGGTTGGTGTGTATGTGCTGGGCT 280
257 TTTCTGATATCTTTTACCAAGCTTTGGCTACAGAACTAGCTACTATGCAATGGGCT 316
281 CCAGCTGGCCCTTTGGCAAGGGCTCTGCAAGGCTGACGGGCTTTGTCTTACACGACA 340
317 TTGACTGGAGATCGGAGATGCCCTTGTAGGATAACTGCGCTAGTGTGTTTACATCAACA 376
341 CCTACGGGGGGTCTACCTCATGCGCTGTGTAGCGTGGACCATTACCCAGCTGTGCTCT 400
377 CATATGAGGTGTGAATTTATGACCTGCTGCTGAGTATTTGACCGCTTCACTGTGCTGTC 436
401 GTGCCACTGGGCGCGCGCTCCGACGCGCTGGCGCGCGAGGCTGGTGTGCTGGGCGCA 460
437 ACCCTGTACGCTACAAACAGATAAAAGGATTAACATGCAAAAGCGGTGTCATATTTG 496
461 TCTGGACCTTGGTGTCTGTCAGACGATGCCCTTGTCTGTGATGCCCATGACCAAGCGCG 520
497 TCTGGATTCTAGTATTTGTGTCAGACACTCCCACTCTCTCATCAACCTTATGTCAAGCAG 556
521 TGTGGGCAAGCTGGCTGCTATGAGTACAGCAGCATGGAGTCAGTCTCGGCTGCCCTCC 580
557 AGGCTGAAGGATTACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 616
581 TCATGCTCTGGTGGCTTTGGCCATTTGGCTTCTGTGGGCGAGTGGGATCATCTGCTCT 640
617 GGATTTCTGCTTGGGCGATGTTTCTAGGATATGATCTTCCACTTAAATCAATTTCTCATCT 676
641 GCTATATGAAGATCACCTGGAAGCTGTGACGACAGCTCGGAGGAGCCAGTGAACAGCA 700
677 GCTATTCTCAGATCTGCTGCAAACTCTTCAAGACTGCAAAACAAACCCACTCACTGAGA 736
701 GGAAGAGCCGCACTGGCGAGGCTGCTGCTTACGCTGCTGCTGCTGCTGCTGCTGCTG 760
737 AATCTGGTGTAAACAAAAGGCTCTCAACACAATTTATTTATTTATTTATTTATTTATTT 796
761 TCTGCTTACGCGCTTACCACTCAACATCAAGCAGTTTCAATGCGGAGAGGATGCTCCACC 820
797 TCTGTTTACACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 856
821 TGCCATCC-----TGTGCGGAGCGGAGGCTTTCTTCTTCTGCTCCCTTACGCGCACCGTGG 874
857 CTAAATTTCTGGAATGTAGCCAAAGACATTCGTTCCAGATTTCTCTGCACTTTTACAGTAT 916
875 CCCTCATGACATCACTGAGGATTTACCCCAATCATTTTACTTCTTTTGTGATCCACCCATT 934
917 GCCTGATGAATTCATTTGCTGATGAGGACCTTTTATCTACTTCTTTTGTGATGAAGGCT 976

QY		935	ACAGAAATGCTCCTCGGCATTTTTAAAGTCCAAAGGGTCTTCCTCCCTCCTCCCTCCCTCCCT	938
DB		977	ATAAGAGAAGAGGTATTAGAGATGCTGAACGGCAAGTCAGTGATCATGATTTCTAGTGCTG	1036
QY		995	CCTCCTCACCCCCCAGGAAAAAGCTTCTTCAGAAAACACA	1032
DB		1037	TGAAGTCAGCCCCTGAAGAAAATTCACGTGAATGACA	1074
RESULT 10 US-10-247-671-39 ; Sequence 39, Application US/10247671 ; Publication No. US20030194721A1 ; GENERAL INFORMATION: ; APPLICANT: Shiffman, Thomas ; APPLICANT: Porter, Gordon, J. ; APPLICANT: Kaser, Matthew R. ; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS ; FILE REFERENCE: PA-0050 US ; CURRENT APPLICATION NUMBER: US/10/247,671 ; CURRENT FILING DATE: 2002-09-18 ; PRIOR APPLICATION NUMBER: 60/323,784 ; PRIOR FILING DATE: 2001-09-19 ; NUMBER OF SEQ ID NOS: 186 ; SOFTWARE: PERL Program ; SEQ ID NO 39 ; LENGTH: 2477 ; TYPE: DNA ; ORGANISM: Homo sapiens ; FEATURE: ; NAME/KEY: misc_feature ; OTHER INFORMATION: Incyte ID No. US20030194721A1 057385.565 ; FEATURE: ; NAME/KEY: unsure ; LOCATION: 2161 ; OTHER INFORMATION: a, t, c, g, or other US-10-247-671-39				
Query Match 16.7%; Score 184.8; DB 14; Length 2477; Best Local Similarity 50.6%; Pred. No. 7.1e-43; Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;				
QY		101	CTCTGTCTTACACAGCCCTCTCGTCTTCAAGTCGCCCTGGGAAACATCTTGGCCCTTTCGCC	160
DB		548	CTCTGATTAACAGCTCTCTTCAATATGGGCTGCTGGAAACTTACTAGCCITGGTGG	607
QY		161	TTACCTGTCAAAAGACGAGGAAGATCAACTGCACAGSCATCTACCTGGTGCACTGGCTG	220
DB		608	TCAATGTTCAAAACAGGAAAAAAAAAACCTCTACCACCTCTATTCAACAAAATTTGGTTA	667
QY		221	TGCTGACCTGCTGTTCACCGTGCCTTACCGGAAGGGTGGTGTATATGTGGCT	280
DB		668	TTTTTGATATACTTTTTTACACCGCTTTGCCTACGATAGCTACTATGCAATGGGCT	727
QY		281	CCAGCTGGCCCTTTCCGCAAGGGGCTCTGCAGGCTGACGGGGTTFTGCTCTACACCGACA	340
DB		726	TTGACTTGGAGATCGAGATGCCTGTGTAGTAGTAACCTGGCTAGTGTATATACATCAACA	787
QY		341	CCTAGGGGGGTCTACCTCATGCGCTGTGACGCTGGACCATTAACCGAGCTGTGCT	400
DB		788	CATATGACGGTGTGAACTTTATGACCTGCTGTAGTATGACCGCTTCATTTGCTGTGGTGC	847
QY		401	GTCGCCACTCTGGGGCCCCGGCCCTCCGCAAGGCTGGCCGCGAGCTGGTCTGCGTGGCCA	460
DB		848	ACCCCTCTACGCTACAACAAGATAAAAAGGATTGAACATGCAAAAGGCGTGTGCATATTG	907
QY		461	TCTGGACCTTGCTGTCTGTGCAGACGATGCCCTTGTCTTTGATGCCCATGACCCAGCGC	520
DB		908	TCTGGATTTCTAGTATTTGCTCAGACACTCCCACTCTCTCATCAACCCATATGTCAAGACG	967
QY		521	TGTTGGGCAAGCTGGCCTGCAATGAGTACAGCAGCATGGAGTCAGTCTCTGGGCTGCCCC	580

968	AGCTGAAGAGTTACATGCATGGAGTATCCAACTTTGAAGAACTAAATCTCTCCCT	1027
581	TCATGCTCTGTGTGGCTTTGGCATTTGGCTTTGTGTGGCCAGTGGGGATCATCTCTCCT	640
1028	GGATTCTGCTTGGGGCATGTTTCATAGGATATGTACTTCCACTTATAATCATCTCATCT	1087
641	GCTATATGAAGATCACTCGAAGCTGTGCAGACACAGCTCGGAGAGACCAGTGACCAGCA	700
1088	GCTATTCTCAGATCTGCTGCAAACTCTTCAGAACTGCCAAACCAAAACCCACTCACTGAGA	1147
701	GGAAAGGACGCCACTGGCGAGGCTGCTGCTTACGCTGCTGATGCTGTGGCGTGGTGG	760
1148	AATCTGCTGTAACAAAAAGGCTCTCAACACAATTAATCTTATTTGTTGTTTGTTC	1207
761	TCGTCTTACGCCCTACCACTCAACATCAAGCAGTTTCATGCGAGAGGATGCTCCACC	820
1208	TCGTGTTACACCTTACCATTTGCAATTTTCAACATATGATTAAAGACTCGTTTCT	1267
821	TGCCATCC-----TGTGCGGAGCGAGGCTTCTTACTGTCCCTTCAGGCCACCGTGG	874
1268	CTAATTCTCTGGAATGTAGCAAGACATTCGTTCCAGATTTCTCTGCACTTTACAGTAT	1327
875	CCCTCATGAACATGAACCTGTGGCATTACCCCAATCATTTACTTTTGTGCATCCCAATT	934
1328	GCCTGATGAACTTCAATTGCTGTCAGACCCCTTTTATCTACTTCTTTGCGATGAAAGGGT	1387
935	ACAGGAAATGGCTCCTGGGCAATTTAAAGCTCAAGGGTCTTCCCTCCTCTCTCCTCCT	994
1388	ATAAGGAAAGGTTATGAGATGCTGCGAAGCGGCAATGATGATGATTTCTAGTGCCTG	1447
995	CTCTCTCCACCCAGGAAAGCTTCTTCAGAAACACCA	1032
1448	TGAAGTCAGCCCTGAGAGAAATTCACGTGAATGACA	1485

RESULT 11
US-10-251-385-205
; Sequence 205, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
; TITLE OF INVENTION: Protein-Coupled
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/10/251,385
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US/09/170,496
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 205
; LENGTH: 1086
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-251-385-205

	Query Match	16.6%	Score 183.2	DB 14	Length 1086
	Best Local Similarity	50.5%	Pred. No. 1.4e-42		
	Matches 474	Conservative 0	Mismatches 458	Indels 6	Gaps 1
QY	101	CTCTGTCTTACACAGCCCTCTGTGCTTTCAGTGCCTTGGGAAACATCCTTGCCTTTGCC	160		
DB	104	CTCTGCATTACAGGCTCGTCTTATCATCTGGGCTGTGGGAACTTACTAGCCCTTGGTCG	163		
QY	161	TTACCTCTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGGTG	220		
DB	164	TCATGTGTTCAAAACAGGAAAAAATCAACTCTATCACCCCTTATTCAACAAATTTGGTGA	223		
QY	221	TGTCGTGACCTGCTCTTTCACCGTGGCCCTTACCGGGAAGGGTGGTGTGTTATGTGTGGGCT	280		

Db 224 TTTCTGATATACCTTTTACAGGCTTTGCTACAGAAAGCTACTATGCAATGGCT 283
Qy 281 CCAGCTGGCTTTTCGGCAAGGGCTCTGAGGCTGACGGGCTTTGCTCTACACGACA 340
Db 284 TTGACTGAGAAATCGGAGATCCCTTGTGTAGATAAATCGGCTAGTGTTTATACAA 343
Qy 341 CCTAGGGGGGTCTACTCTATGCTGTGTGAGGCTGAGCAATACCCAGCTGTGTCT 400
Db 344 CATATGCAAGGTGTGAACCTTTATGACCTTGCCTGAGTATTTGACCGCTTCA 403
Qy 401 GTGCCACTGGGGCCCGCTCCGCAAGGCTGCGCGCGAGGCTGGTCTGGGTGGCCA 460
Db 404 ACCCTCTAGCTACAAAGATATAAAGGATTTGAACATGCAAAAGGCTGTGCATATT 463
Qy 461 TCTGACCTTGGTCTGTGAGAGATGCTTGTGCTTGTGATGCCCATGACCAAGCGC 520
Db 464 TCTGGAATCTAGTATTGCTCAGACACTCCCACTCCCTCATCAACCTTATGTCAA 523
Qy 521 TGGTGGCAAGCTGGCTTGCATGAGTACAGCAGCATGAGTCTGCTCGGCTGCCCC 580
Db 524 AGGCTGAAAGATTTACATGATGAGTATCCAAACTTTGAAGAACTAAATCTCTCC 583
Qy 581 TCATGGTCTGGTGGCTTTGGCAATGCTTCTGTGGCCAGTGGGATCATCTGTCTCT 640
Db 584 GGAATCTGCTTGGGCGATGTTTCAATAGGATATGTAATCTTCACTTATAATCAT 643
Qy 641 GCTATATGAAGATCACTTGAAGCTGTGAGCAGCATGCTCGGAGGACCCAGTGA 700
Db 644 GCTATCTCAGATCTGCTGCAACTTTCAGAACTGCGCAACAAACCACTCACTGA 703
Qy 701 GGAAGGAGCGCACTGGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 760
Db 704 AATCTGGTGTAAACAAAAGGCTTAAACACAAATTAATCTTATTTGTTGTTGTC 763
Qy 761 TCTGCTTCCAGCCCTTACCACTCAACATCAAGCAGTTCATGCGAGAGGATGCTCA 820
Db 764 TCTGTTTACACCTTACCAATGTTGCAATTTTCAACATATGATTAAGAGCTTCT 823
Qy 821 TGCCATCC-----TGTGCGAGCGAGGCTTTCTTACTGTCCTTCCAGCCACCGT 874
Db 824 CTAAATTTCTGGAATGTAGCAAGACATTCCTCCAGATTTCTGCACTTTACAGTAT 883
Qy 875 CCTCATGACATGACTGTGCAATACCCCAATCATTTCTTCTTCTGCACTTCCACCAT 934
Db 884 GCTGATGAATTCATTTGCTGCTGAGGCTTTTATCTTCTTCTGCAATGAAGGCT 943
Qy 935 ACAGAAATGGCTCTCTGGGCAATTTTAAAGCTCAAGGCTTCTTCTCTCTCTCTCT 994
Db 944 ATAAGAGAAAGTATGAGATGCTGAAACGCAAGTCAAGTATCGATTTCTAGTGTG 1003
Qy 995 CTTCTCCACCCAGAAAGCTTCTTCAAGAACCA 1032
Db 1004 TGAAGTCAAGCCCTGAAGAAATTCACGTGAATGACA 1041

RESULT 12

US-10-225-567A-111
; Sequence 111, Application US/10225567A
; Publication No. US2003013798A1
; GENERAL INFORMATION:
; APPLICANT: LifeSpan Biosciences
; APPLICANT: Brown, Joseph P.
; APPLICANT: Burmer, Glenna C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225,567A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 2292
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 111
; LENGTH: 1638
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-225-567A-111
Query Match 16.6%; Score 183.2; DB 14; Length 1638;
Best Local Similarity 50.5%; Pred. No. 1.7e-42;
Matches 474; Conservative 0; Mismatches 458; Indels 6; Gaps 1;
Qy 101 CTCTGTTCTACACAGCCCTCTGCTTTCAGTCCCTGGAAACATCTTGGCCCTTGGC 160
Db 137 CTCTGCAATACAGCTCGTCTTCATCATTTGGGCTCGTGGAAACATCTAGCTTGG 196
Qy 161 TTACCTGTCAAAAGACAGGAAGATCAACTGCACAGGCACTTACCTGGTGCACCTG 220
Db 197 TCATTTCTCAAAACAGGAATAAATCAACTCTACCACTCTATTCAACAAATTTGG 256
Qy 221 TGTCTGACCTGTCTTACCTGCGCTTACCGGGAAGGTGTGTGTATGTGCTGGGCT 280
Db 257 TTTCTGATATACCTTTTACCACCGCTTTGCCCTTACGAATAGCCCTACTATGCA 316
Qy 281 CCAGCTGGGCTTTTGGCAAGGGCTCTGCAAGCTGACGGGCTTTGTGCTCTACACG 340
Db 317 TTGACTTGGAGATCGAGATGCTTGTGTAGATAAATCGCTAGTGTATTACATCA 376
Qy 341 CCTACGGGGGGTCTTACCTCATGCGCTGTGTGAGCGTGGAACATACCCAGCTGT 400
Db 377 CATATCAGGTGTGAATTTATGACCTGCTGAGTATTGACCGCTTTCATGCTGTG 436
Qy 401 GTGCCCACTGGGCGCGCTCCGCAAGGGCTTCCGCGCGCGCGAGGCTGTGCTGCT 460
Db 437 ACCCTCTAGCTACAAAGATAAAGATTTGAACATGCAAAAGCGGTGCAATTTG 496
Qy 461 TGTGACCTTGTGTGCTGCAAGCAATGCTTGTCTTGTGTGATGCCCATGACCAAGC 520
Db 497 TCTGGATTTCTAGTATTGCTTCAACACTCCCTCATCAACCCCTATGTCAAGACAG 556
Qy 521 TGTGTGGCAAGCTGCGCTGCTGAGTACAGCAGCATGAGTCAAGTCAAGTCTCGG 580
Db 557 AGGCTGAAAGGATTAATGATGAGTATCCAACTTTGAAGAACTAAATCTTCTCC 616
Qy 581 TCATGTCTCTGTGGCTTTTGCATTTGGCTTTCTGTGGGCGAGTGGGATCATCTGT 640
Db 617 GGATTTCTGTGGGCTATTTTCTAGGATATGTTCTTCCACTTATAATCAATCTCA 676
Qy 641 GCTATATGAAGATCACTGGAAGCTGTGAGCAGCTCGGGAGGACCCAGTGAACGA 700
Db 677 GCTATTTCTAGATCTGCTGCAAACTCTTCAAGAACTGCCAAACAAACCACTCA 736
Qy 701 GGAAGGACCCACTGCGAGGCTGCTGCTTACGCTGTGTGATGCTGTGGCGCTGGT 760
Db 737 AATCTGGTGAACAAAGGCTCTCAACACATTAATCTTATTGTTGTTGTTGTTT 796
Qy 761 TCTGTTTCAAGCCCTTACCACTCAATCAAGAGTTCAAGAGGATGCTCCACC 820
Db 797 TCTGTTTCAAGCCCTTACCACTTGTGCAATTAATCAACATATGATTAAGAGCTTC 856
Qy 821 TGCCATCC-----TGTGCGAGCGAGGCTTTTCTTACTGTCCCTTCAGGCCACCG 874
Db 857 CTAAATTTCTGGAATGTAGCAAGACATTCGTTCCAGATTTCTCTGCATTTACAG 916
Qy 875 CCTCATGAACATGAATGTGGCAATACCCCAATCAATCTTCTTGTGCAATCCACCAT 934
Db 917 GCCTGATGAATCTCAATTTGTGCAATGACCCCTTTTATCTTCTTGTGATGAAGGG 976
Qy 935 ACAGGAATGCTCTCTGGGCAATTTTAAAGCTCAAGGCTTCTTCTCTCTCTCTCTCT 994
Db 977 ATAAGAAAGGTTATGAGATGCTGAAACGGCAAGTCACTGTATCGATTTCTAGTCT 1036
Qy 995 CCTCTCCACCCAGGAAAGCTTCTTCAAGAAACCA 1032
Db 1037 TGAAGTCAAGCCCTGAAGAAATTCAGTGAATGACA 1074

```

RESULT 15
US-09-918-995-10498
; Sequence 10498, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918.995

```

```
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10498
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Hmo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(548)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-10498

Query Match      11.2%; Score 123.4; DB 10; Length 548;
Best Local Similarity 52.8%; Pred. No. 2.5e-25;
Matches 262; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

QY 167 GTCAAAGACGAGGAGATCAACTGCACGAGGAGTCTACCTGGTGCACCTGGCTGTCTG 226
DB 53 GNTANACAGGAGAAATCAACTTACACCCCTTATTCAACAAATTTGGGTATTCTG 112

QY 227 ACCTGCTGTTCACCGTGGCCCTTACCGGAGGGTGTGTATTGTGCTGGCTCCAGCT 286
DB 113 ATATATTTTTTACCAACCGCTTTGCCTACACGAATAGCCTACTATGCAATGGCTTTGACT 172

QY 287 GGCCTTTGGCAAGGGGCTCTGCAGGCTGACGGGTTTGTGCTCTACACGACACCTACG 346
DB 173 GGAGAATCGGAGACGCCCTTGGGTAGGATAACTGGGCTAGTGTTTTACATCAACACATATG 232

QY 347 GGGGGGTCTACCTCATGGCTGTGTGAGCGTGGACCATTAACCCAGCTGTGTGTGCCC 406
DB 233 CAGGTGTGACTTTATGACCTGCTGCTGATTTGACCGCTTTCATGCTGTGTGCACCTC 292

QY 407 ACTGGGGCCCGCCCTCCGACGCGTGGCGGCGCCAGCGTGGTCTGGTGGCCATCTGGA 466
DB 293 TACGCTACAAACAGATAAAAGGATTGAACATGCAAAAGGCGTGTGCATATTGTCTGGA 352

QY 467 CCTTGGTGTCTGTGCAGAGGATGCCCTTGTCTTGTATGCCATGACCAAGCGCTGGTGG 526
DB 353 TTCTAGTATTGCTCAGACACTCCCACTCTCATCAACCTATGTCAAGCAGGAGGCTG 412

QY 527 GCAAGCTGGCCTGSCATGGAGTACAGCAGCATGGAGTCACTCTCGGGGTGCCCTCATGG 586
DB 413 AAAGGATTACATGCATGGAGTATCCAAACTTTGAAGAAACTAAATCTCTCCCTGGATTC 472

QY 587 TCCTGTGGCCTTTGGCATTTGCTTCTGTGGCCAGTGGGATCATCTCTGCTCTGCTATA 646
DB 473 TGGTTGGGCATGTTTTCATAGGATATGTACTTCCACTTATAATCATTCATCTCTAT 532

QY 647 TGAAGATCACCTGGAA 662
DB 533 CTCAGATCTGCTGCA 548
```

Search completed: February 21, 2004, 21:36:27
Job time : 474 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 19:11:15 ; Search time 2793 seconds
(without alignments)
11803.741 Million cell updates/sec

Title: US-10-041-615-33
Perfect score: 1104
Sequence: 1 atgataaactaggccctta.....gtgtcattgtgttatgact 1104

Scoring table: IDENTITY.NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_tod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*
29: gb_gss2.*

5 160.2 14.5 979 9 AL532537
6 160.2 14.5 1201 13 BX417963
7 159.4 14.4 636 14 CF359817
8 159.4 14.4 1201 9 AL551903
9 155.6 14.1 1088 12 BM549799
10 155.2 14.1 715 13 BY765663
11 151.2 13.7 963 13 BY745873
12 150.6 13.6 1201 13 BX427288
13 148 13.4 594 9 AI659965
14 143.4 13.0 662 13 BY729793
15 140.4 12.7 861 12 BY768868
16 139.4 12.6 660 13 BQ563908
17 139 12.6 679 13 BY751377
18 138.4 12.5 611 14 CF913347
19 138.4 12.5 682 13 BY750933
20 137.2 12.4 933 13 BX453641
21 133.4 12.1 778 14 CD468929
22 133 12.0 614 10 B8637586
23 124 11.2 678 14 CD687357
24 124 11.2 695 14 CD686410
25 124 11.2 749 12 B1821236
26 122.4 11.1 686 14 CD708189
27 121.8 11.0 1201 9 AL575476
28 121.6 11.0 671 14 CD640122
29 119.6 10.8 682 13 BY750497
30 119 10.8 820 14 CF181257
31 118.4 10.7 828 10 B8636985
32 118 10.7 574 12 B1009654
33 117.6 10.7 673 14 CD465254
34 115.6 10.5 643 14 CF249695
35 115.4 10.5 673 12 B8547923
36 115 10.4 631 14 CD470715
37 113.6 10.3 879 12 BQ536922
38 112.4 10.2 660 12 B1394962
39 111.4 10.1 539 14 CA575605
40 110.6 10.0 532 14 CA576608
41 108.2 9.8 789 14 CD521415
42 107.6 9.7 679 10 B8663009
43 106.4 9.6 1019 12 B1517811
44 105.8 9.6 4022 11 AK043873
45 102.6 9.3 700 12 BG483806

ALIGNMENTS

RESULT 1
AK087951
LOCUS
DEFINITION Mus musculus 2 days neonate thymic cells cDNA, RIKEN
full-length enriched library, clone:R33001G12 product:EBV-INDUCED
G PROTEIN-COUPLED RECEPTOR 2 (EBI2) homolog [Homo sapiens], full
insert sequence.

ACCESSION AK087951
VERSION AK087951.1 GI:26104619
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
AUTHORS Carninci P. and Hayashizaki Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 309, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE 2
AUTHORS Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K.,
Itoh M., Kono H., Okazaki Y., Muramatsu M. and Hayashizaki Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	211.2	19.1	1496	11 AK087951	AK087951 Mus muscu
2	180.6	16.1	631	14 CB499474	CB499474 ssalkhco1
3	169.6	15.4	1097	9 AL547762	AL547762 AL547762
4	168.4	15.3	1171	13 BX356015	BX356015 BX356015

11042159
 3 SHIBATA,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
 Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
 Sumioto,R., Matsumoto,R., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
 Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,J.,
 Okazaki,Y., Muranatsu,M., Inoue,Y., Kira,A., and Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000

```
RESULT 2
CB499474          631 bp      mRNA      linear      EST 16-MAY-2003
LOCUS             ssalkh013018 kidney salmo salar cdna, mRNA sequence.
DEFINITION       CB499474
ACCESSION        CB499474
VERSION          CB499474.1 GI:29310700
KEYWORDS         EST.
SOURCE           Salmo salar (Atlantic salmon)
ORGANISM         Salmo salar
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
1 (bases 1 to 631)
GRASP Consortium, Davidson, W.S., Koop, B.F. and
http://web.uvic.ca/cbr/grasp/
A survey of Salmo salar transcripts from high complexity cdna
libraries
Unpublished (2002)
Contact: Koop BF
Centre for Biomedical Research
University of Victoria
PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
Tel: 250 472 4067
Fax: 250 472 4075
Email: bkoo@uvic.ca
Centre for Biomedical Research, University of Victoria cdna
preparation and sequencing: Roberto Alberto, Marianne
Bestz-Sargent, Maura Busby, Peter Hunt, Linda McKinnel, BF Koop.
bioinformatics: Gordon D Brown.
Location/Qualifiers
1. .631
/organism="Salmo salar"
/mol_type="mRNA"
/strain="McConnell"
/db_xref="taxon:8030"
/clone_lib="kidney"
/note="Vector: pBlueScriptII SK+; Library Creator: Matthew
L Rise ; Atlantic salmon tissue contributors: Carlo Biagi,
Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon
Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton,
B.C.), Rachel Roper (University of Victoria)"
FEATURES             source
Query Match          16.4%; Score 180.6; DB 14; Length 631;
Best Local Similarity 56.6%; Pred. No. 2.5e-27;
Matches 355; Conservative 0; Mismatches 269; Indels 3; Gaps 1;
QY 178 AGGAGATCAACTGCACAGGCATCTACCTGGTGACCTGGCTGTGTCTGACCTGCTGTC 237
DB 3 AAGAGATTAACTCCACACGCTTACTCTGCCACCTGGTCACTCCGACATCTGTTTC 62
QY 238 ACCGTGGCCTTACCGGGAAGGTGTGTGTGTATGTGTGGCTCGGCTCCAGCTGGCCTTGGC 297
DB 63 ACACCTGTCACTACCGTTACGGTAATCTACTACGCCCTGGGCTTCCACTGGCCCTGGGG 122
QY 298 AAGGGCTCTGAGGCTGACGGCTTTGTGCTCTACACGACCTACGGGGGTCTAC 357
DB 123 GAGACCTGTGTAGATCTGGGGTGATCTTCTACATCATACTTAAGCGGGGTCAAC 182
QY 358 CTCATGGCTGTGTAGCTGACCATTAACCATGCTGTGTGTGTGCCCATCTGGGGCCG 417
DB 183 TTTCATGACCTGCTCAGTGTGACCGGTTCAATTGCTGTGCTTGGCTGGCATTCGCA 242
QY 418 CGCTCCGACAGCTGGCGCGCGAGCTGTGTGTGTGGCCATCTGGACCTTGTGTGTG 477
DB 243 CGCTTCGTAAGTCTTCCAATGTCCGTTACATTGGCTGGGGTGTGTGTGTGTCTTA 302
QY 478 CTGCAGACGATGCCCTTGTCTTGTATGCCCATGAC---CAAGCCGCTGTGGGCAAGGTG 534
DB 303 ATGCAAACTCTGCCCTCTCTCCATGCCCATGACTTAACGAGAACTGATGGCTTCATC 362
QY 535 GCTGTGATGAGTACAGCAGCATGAGTCAGTCTCGGCTCGGCTCGGCTCTGCTGTGTG 594
```

```
DB 363 ACCTGTATGGAGTATCCCAACTTCGAAACGGGTGCTCAACATCTCTATATCTCTGATTGGC 422
QY 595 GCCTTTGCCATTGGCTTCTGTGGGCACTGGGATCATCTCTCTCTATATGAAGATC 654
DB 423 GCGGTATTCTTAGGTACGGAGTCCCGGTGGTGGACCATCTCTGTGTCTACTTCACTG 482
QY 655 ACCTGGAAGCTGTGCAGCACAGCTCGGAGGACCCAGTGACCCAGCAGGAAGAGCCAC 714
DB 483 TCGTGTAAACTCCGCTCGCGCCCAAGGCCAATCAGCTGACGACAAGTCGGGGGCGAGC 542
QY 715 TGGCAGAGGTGCTCTGCTGCTAGCTGTGTGTGTGTGGCGGTGGTGTCTCTCAGCCCC 774
DB 543 CAAAAGCCATCGGGGTGATCTGCTGCGCTCTCCCTGGTGTGTGTCTGTTTTCAGCCCC 602
QY 775 TACCACCTCAACATCAAGCAGTTCATG 801
DB 603 TATCATCGACTCTCTCAATACATG 629

AL547762          1097 bp      mRNA      linear      EST 31-MAY-2003
LOCUS             AL547762
DEFINITION       AL547762 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cdna
clone CS0D1017YN05 5-PRIME, mRNA sequence.
ACCESSION        AL547762 GI:31269591
VERSION          AL547762.2
KEYWORDS         EST.
SOURCE           Homo sapiens (human)
ORGANISM         Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1097)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:12882129.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6994.r For
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0D1017CG03QPL&cluster=6994.r. Contact :
Peng Liang Email: fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0D1017CG03QPL.
Location/Qualifiers
1. .1097
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
FEATURES             source
Query Match          15.4%; Score 169.6; DB 9; Length 1097;
Best Local Similarity 49.7%; Pred. No. 5.9e-25;
Matches 434; Conservative 10; Mismatches 424; Indels 6; Gaps 1;
QY 101 CTCTGTCTACACGCCCTCTGCTTTCAGTCCCTGGGAAACATCTCTGCTTGGCCTTGGC 160
DB 222 CTCTGCATTACAGCTCTGCTTTCATTTGGGTCTGGGAACTTACTAGCTTGGTGG 281
QY 161 TTACTCTCAAAAGAGCAGGAAGATCAATGTCACAGCATCTCTGCTGGTGCACCTGGCTG 220
DB 282 TCATTGTTCAAACAGGAGAAAAAATCACTCTACCCCTCTATTCAACAAATTTGGTGA 341
```

221 TGTCTGACCTGCTGTTTACACGCTGGCCCTTACCGGGAAGGGTGGTGTATGCTGGGCT 280
 Db TTTCTGATATATCTTTTACACCGCTTGTGCTACAGATAGCTACTATGCAATGGGCT 401
 281 CCAGCTGGCCTTTGGGCAAGGGCTCTGAGAGCTGACGGGTTTGTGCTCTACACCGACA 340
 Db TTGACTGAGAAATCGGAGATGCGCTTGTGTAGGATAACTGCGCTAGTGTGTTTACATCAACA 461
 341 CCTACGGGGGGGCTTACCTCATGCGCTGTGTGAGCGTGGACCAATACCAGAGTGTGGTCT 400
 Db CATATGACAGGTGTGAATCTTATGACCTGCTGAGTATGACCGCTTATGCTGTGGTGC 521
 401 GTGCGCACTGGGGCCCGGCTCGGACGCTGCGCGCGCGGCGGCTGCTGCTGGGCA 460
 Db ACCCTCTACGCTACAAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAG 581
 461 TCTGACCTGCTGCTGCTGACAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 520
 Db TCTGGATTCTAGTATTTGCTCAGACACTCCACTCTCTCATCAACCTATGTCAGAGCAGG 641
 521 TGGTGGGCAAGCTGGCTGATGAGGATGACAGAGATGAGTCACTCTCGGGCTGGCC 580
 Db AGGCTGAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAG 701
 581 TCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 640
 Db GGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 761
 641 GCTATATGAGATCAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCT 700
 Db GCTATCTCAGATCTGCTGAGAACTCTTCAAGACTGCTGAGAACTGCTGAGAACTGCTGAG 821
 701 GGAAGAGAGCTGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 760
 Db AATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 881
 761 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 820
 Db TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 941
 821 TGCATCT 874
 942 CTAAATCTCTGGAATGAGCAAGAAATTCGTTCCAGATTTCTCTGCACTTAAAGAT 1001
 875 CCTCATGACATGACCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 934
 Db CCTGATGATCTTAAATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1061
 935 ACAGAAATGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 968
 Db AAAAAAAAGGTATTAAGGATGYSRAACGGAA 1095

RESULT 4

BX356015

LOCUS

DEFINITION

BX356015 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA

clone CSOD1007Y101 5-PRIME, mRNA sequence.

ACCESSION

BX356015

VERSION

BX356015.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1171)

Li, W.B., Gruber, C., Jesse, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqrefgenoscope.cns.fr, Web: www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 6994.1. For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CSOD1007AE01QP2&cluster=6994.1. Contact :
 Feng Liang Email: fliang@life.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CSOD1007AE01QP2.

Location/Qualifiers
 1. 1171
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CSOD1007Y101"
 /issue_type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with NotI and cloned into the NotI and EcoRV
 sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES

source

ORIGIN

Query Match 15.3%; Score 168.4; DB 13; Length 1171;
 Best Local Similarity 50.5%; Pred. No. 1.1e-24;
 Matches 420; Conservative 7; Mismatches 398; Indels 7; Gaps 1;
 101 CTCTGTTCTACACAGCCCTCTCTGCTTTCAGTGCCTGGGAAACATCCTTSCCTTGGCC 160
 Db CTCTGATTACAGCTCGTCTTCATCATTTGGGCTCGTGGAACTTACTAGCTTGGTCG 246
 161 TTACCTGTCTCAAGAGCAGGAAGATCAACTGCACAGGATCTACCTGTGCACCTGGCTG 220
 Db TCATTTCTCAAAACAGGAAAAAATCAACTCTACACCTCTATTCAACAAATTTGGTGA 306
 221 TGTCTGACCTGCTGTTTACCGTGGCTTACCGGAGGGTGTGTTATGCTGGCT 280
 Db TTCTCTGATATATTTTACACCGCTTGGCTTACAGAACTAGCTTACTATGCAATGGGCT 366
 281 CCAGCTGGCTTTCGCGAAGGGCTCTGACGGCTGACGGCTGTTGTGCTCTACACCGACA 340
 Db TTGACTGGAGATCGGAGATGCTTGTGTAGATAACTGCGTAGTGTTCATCAACA 426
 341 CCTACGGGGGCTTACCTCATGCTGCTGAGGCTGAGGCTGACCAATACCAGCTGGTCT 400
 Db CATATGAGGTGTGAACCTTATGACCTGCTGAGTATGACCGCTTCAATCTGTGGTGC 486
 401 GTGCCCACTGGGGCCCGGCTCTCCGACGGCTGGCGCGCCAGGCTGCTCTGCGTGGCCA 460
 Db ACCCTCTACGCTACAAAGATTAAGAAAGATTAAGAAAGGCTGTGCATATTG 546
 461 TCTGACCTTGTGCTGCTGAGACGATGCTTCTTCTTCTGATGCCATGACCAAGCGC 520
 Db TCTGATTTCTAGTATTTGCTGACACCTCCCACTCTCTCATCAACCTATGTCAAGCAGG 606
 521 TGTGGGCAAGCTGGCTGCTGATGAGTACAGCAGCATGGAGTCACTCTCGGGCTGGCCC 580
 Db AGGCTGAAGAGTATACATGATGAGTATCCAACTTTTGAAGAACTAAATCTCTTCCCT 666
 581 TCATGGTCTGCTGGGCTTGGCCATTTGCTTGTGGCCAGTGGGATCATCTGCTCT 640
 Db GGATTTCTGCTTGGGCTGTTTCTAGGATATGATTTCCACTTATATCAATTTCTATCT 726
 641 GCTATATGAAGATCACTCGAAGCTGTGCAGCAGCTCGGAGGAGCCAGTACCAAGCA 700
 Db GCTATTTCTCAGATCTGTGCAAACTCTTTCAGAACTGTCGCAAAACAAACCTCACTGAGA 786
 701 GGAAGGAGCCACCTGGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 760
 Db AATCTGGTGTAAACAAAGGCTCTCAACAATTTCTTATTATTGTTGTTGTTTCT 846
 761 TCTGCTTACGCCCCCTACCACTTCAACAGCTTCAAGCAGTTTCATGGCGAGAGGATGCTCCACC 820

Email: seqrefgenoscope.cns.fr, Web: www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 6994.1. For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CSOD1007AE01QP2&cluster=6994.1. Contact :
 Feng Liang Email: fliang@life.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CSOD1007AE01QP2.

Location/Qualifiers
 1. 1171
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CSOD1007Y101"
 /issue_type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with NotI and cloned into the NotI and EcoRV
 sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES

source

ORIGIN

Query Match 15.3%; Score 168.4; DB 13; Length 1171;
 Best Local Similarity 50.5%; Pred. No. 1.1e-24;
 Matches 420; Conservative 7; Mismatches 398; Indels 7; Gaps 1;
 101 CTCTGTTCTACACAGCCCTCTCTGCTTTCAGTGCCTGGGAAACATCCTTSCCTTGGCC 160
 Db CTCTGATTACAGCTCGTCTTCATCATTTGGGCTCGTGGAACTTACTAGCTTGGTCG 246
 161 TTACCTGTCTCAAGAGCAGGAAGATCAACTGCACAGGATCTACCTGTGCACCTGGCTG 220
 Db TCATTTCTCAAAACAGGAAAAAATCAACTCTACACCTCTATTCAACAAATTTGGTGA 306
 221 TGTCTGACCTGCTGTTTACCGTGGCTTACCGGAGGGTGTGTTATGCTGGCT 280
 Db TTCTCTGATATATTTTACACCGCTTGGCTTACAGAACTAGCTTACTATGCAATGGGCT 366
 281 CCAGCTGGCTTTCGCGAAGGGCTCTGACGGCTGACGGCTGTTGTGCTCTACACCGACA 340
 Db TTGACTGGAGATCGGAGATGCTTGTGTAGATAACTGCGTAGTGTTCATCAACA 426
 341 CCTACGGGGGCTTACCTCATGCTGCTGAGGCTGAGGCTGACCAATACCAGCTGGTCT 400
 Db CATATGAGGTGTGAACCTTATGACCTGCTGAGTATGACCGCTTCAATCTGTGGTGC 486
 401 GTGCCCACTGGGGCCCGGCTCTCCGACGGCTGGCGCGCCAGGCTGCTCTGCGTGGCCA 460
 Db ACCCTCTACGCTACAAAGATTAAGAAAGATTAAGAAAGGCTGTGCATATTG 546
 461 TCTGACCTTGTGCTGCTGAGACGATGCTTCTTCTTCTGATGCCATGACCAAGCGC 520
 Db TCTGATTTCTAGTATTTGCTGACACCTCCCACTCTCTCATCAACCTATGTCAAGCAGG 606
 521 TGTGGGCAAGCTGGCTGCTGATGAGTACAGCAGCATGGAGTCACTCTCGGGCTGGCCC 580
 Db AGGCTGAAGAGTATACATGATGAGTATCCAACTTTTGAAGAACTAAATCTCTTCCCT 666
 581 TCATGGTCTGCTGGGCTTGGCCATTTGCTTGTGGCCAGTGGGATCATCTGCTCT 640
 Db GGATTTCTGCTTGGGCTGTTTCTAGGATATGATTTCCACTTATATCAATTTCTATCT 726
 641 GCTATATGAAGATCACTCGAAGCTGTGCAGCAGCTCGGAGGAGCCAGTACCAAGCA 700
 Db GCTATTTCTCAGATCTGTGCAAACTCTTTCAGAACTGTCGCAAAACAAACCTCACTGAGA 786
 701 GGAAGGAGCCACCTGGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 760
 Db AATCTGGTGTAAACAAAGGCTCTCAACAATTTCTTATTATTGTTGTTTCT 846
 761 TCTGCTTACGCCCCCTACCACTTCAACAGCTTCAAGCAGTTTCATGGCGAGAGGATGCTCCACC 820

```

847 TCTGTTTACACCTTACCATGTTGCAATTATTAACAATATGATTAAGAAGCTGCTTCT 906
821 TGCCATCTGT-----GCGAGCGGAGGCTTCTTACTGTCCTTACGCCACCGTG 873
907 CTAATTTCTTGGAATGAGCAAGACATTCGTTTCAGATTCTCTGCACTTTACAGTA 966
874 GCCCTCATGAACATGAATGCTGGCATACCCCAATCAATTTACTTCTTTGCAT 925
967 TGCCTGATGACATCAATGTTGTCATGGGCCCTTTATATCTATCTATCTATCTAT 1018

RESULT 5
AL532537 979 bp mRNA linear EST 23-MAY-2003
DEFINITION Homo sapiens FETAL LIVER Homo sapiens cDNA clone
CSODM006YC06 5-PRIME, mRNA sequence.
ACCESSION AL532537
VERSION AL532537.2 GI:31070369
KEYWORDS EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 979)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 13, 2001 this sequence version replaced gi:12796030.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6994.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODM006B03QPl&cluster=6994.r. Contact :
Feng Liang Email : fliang@lifetech.com URL : Corporation 1600
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODM006B03QPl.
Location/Qualifiers
1. 979
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODM006YC06"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/note="Organ: Liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN
Query Match 14.5%; Score 160.2; DB 9; Length 979;
Best Local Similarity 51.8%; Pred. No. 5.4e-23;
Matches 363; Conservative 0; Mismatches 338; Indels 0; Gaps 0;

101 CTCCTGTTTACACGCTCTGCTGCTTCAGTGCCTCGGAAACATCTTCCCTTTGCC 160
255 CTCCTGATTACAGCCTCGCTTTCATCATTTGGGCTCGTGGAACTTACTAGCCTTGGTGC 314
161 TTACCTGTCAAGAGCAGGAAGATCAACTGCACAGGCATCTACTGTCGACCTGGCTG 220
315 TCATTGTTCAAAACAGGAAATAATCACTTACACCTCTATTCAAAATTTGTGA 374
221 TGTCTGACCTGCTGTTCACCGTGGCTTACCGGGAGGGTGGTGTGTTATGCTGGGCT 280
375 TTTCTGATATATCTTTTACCACCGCTTTTGCCTTACACGAATAGCCTTACTATCAATGGCT 434
281 CCAGCTGGCCCTTTCGGCAGGGGCTCTGCAGGCTGACGGGCTTGTGCTCTACACGACA 340

```

```

435 TTGACTGGGAATCGGAGATGCTTGTGTAGATAACTGCGTAGTGTTTTACATCAACA 494
341 CTAAGCGGGGGTCTACCTCATGGCTGTGTAGCGTGGACCATTACCCAGCTGTGCT 400
495 CATATGCAAGGTGTGAACCTTTATGACCTGCCCTGAGTAATGACCGCTTCATGCTGGTGC 554
401 GTGCCCACTGGGGCCCGGCGCTCCGCACGGGTGGCGCGCCAGGCTGCTCTGCGTGGCCA 460
555 ACCCTCTACGCTACAACAAGATATAAAGATTGAACATGCAAAAGCGCTGTGATATTG 614
461 TCTGACCTTGTGCTGCTGCAGAGATGCCCTTCTCTTGTATGCCATGACCAAGCGC 520
615 TCTGATTCTTAGTATTGCTCAGACACTCCCACTCTCATCAACCTATGTCAAAAGCAGG 674
521 TGTGCGGCAAGCTGGCTGCATGGAGTACAGCAGCATGGAGTCAGTCCCTCGGGCTGCCCC 580
675 AGGCTGAAGGATTACATGCAATGGAGTATCAAACTTTGAAGAACTAAATCTCTTCCCT 734
581 TCATGCTCTGTGCTGCTTGGCAATGCTGCTGTGGGCGCAGTGGGGATCATCTGTCT 640
735 GGATTCTGCTTGGGCGATGTTTTCATAGGATATGTATCTTCCACTTATAATCATCTCATCT 794
641 GCTATATGAAGATCACTGGAGCTGTGCAGCACAGCTCGGAGGACCCAGTGACAGCA 700
795 GCTATTCTCAGATCTGTGCAAACTTTCAGAACTGCCAAACCAACCACTCATCTGAGA 854
701 GGAAAGGACGCCACTGGGAGGCTGCTTACGCTGCTGATGCTGTGGCGGTGGTGG 760
855 AATCTGCTGTAACAACAAAGGCTCTCAACACAATTAATCTTATTATTGTTGTTGTTTC 914
761 TCTGTTGAGCCCTTACCACCTCAACATCAAGCACTTTCATG 801
915 TCTGTTTACACCTTACCACATGTTGCAATTAATCAACATATG 955

RESULT 6
BX417963 1201 bp mRNA linear EST 13-MAY-2003
LOCUS BX417963 Homo sapiens PLACENTA Homo sapiens cDNA clone CSODE014YH17
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION BX417963
VERSION BX417963.1 GI:30633353
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6994.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODE014YH17&cluster=6994.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODE014YH17.
Location/Qualifiers
1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODE014YH17"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

```

with a NotI-oligo(dN) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN	Query Match	Score	160.2;	DB 13;	Length	1201;
	Best Local Similarity	51.8%;	Pred. No. 5.7e-23;			
	Matches	363;	Conservative	0;	Mismatches	338;
					Indels	0;
					Gaps	0;
QY	101	CTCTGTTCTACAGACCCCTCTGCTTTCAGTGCCTCGGGAACATCTTGCCTTTGCC	160			
Db	211	CTCTGCATTACAGCCCTGCTTTCATCTCATTTGGCTGCTGGGAACCTTACTAGCCTTGTGTCG	270			
QY	161	TTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACTGCTGGTGCACCTGGCTG	220			
Db	271	TCATTTGTTCAAAACAGGAAAAATCACTTACACCCCTCTATTCAACAAATTTGGTGA	330			
QY	221	TGCTGACCTGCTGTTTACCGTGGCTTACCGGGAGGGTGTGTATGTGTGCTGGCT	280			
Db	331	TTTCTGATATATCTTTTACCACCGCTTTGGCTTACAGGAATAGCCTACTATGCAATGGGCT	390			
QY	281	CCAGCTGGCTTTGCGAAGGGCTCTCAGGCTGACGGCTTTGTGCTTACACCGACA	340			
Db	391	TTGACTGGAGATCGAGATGCTTGTGTAGATPACTGGCTAGTGTTTTACATCAACA	450			
QY	341	CCTACGGGGGCTCTACTCTCATGCTGTGTAGGCTGGACCATTAACCCAGCTGTGTCT	400			
Db	451	CATATGCAAGTGTGAATTTATGACCTGCTGATGATTTGACCGCTTCTATTGCTGTGTGTC	510			
QY	401	GTGGCCACTGGGGCCCGCCCTCCGACGGCTGGCGCGCCAGGCTGTGTGCTGGGCCA	460			
Db	511	ACCTCTACGCTACAAACAGATAAAAGATTGAACATGCAAAAGGCGTGTGCATTTTG	570			
QY	461	TCTGGACCTTGTGTGCTGCAGAGATGCCCTTCTGCTTGTATGCCCAATGACCAAGCGC	520			
Db	571	TCTGGATTTCTAGTATTGTGTCAGACACTCCCACTCTCATCAACCTATGTCAAAAGCAGG	630			
QY	521	TGGTGGGGAAGTGGCTGCATGAGTACAGACAGATGAGTCACTCCTCGGGTGGCC	580			
Db	631	AGGCTGAAGAGATTACATGCATGGAGATTCACAACTTTGAAGAACTAAATCTCTTCCCT	690			
QY	581	TCATGCTCTGCTGGCTTTGCAATTTGCTTCTGTGGCCAGTGGGGATCATCTCTGCTCT	640			
Db	691	GGATTTCTGCTTGGGGATGTTTCATAGGATATGATCTTCCACTTATATCATTTCTCATCT	750			
QY	641	GCTATATGAAGATCACTGGAGCTGTGCAGACAGCTCGGAGACCCAGTGCACCA	700			
Db	751	GCTATTCTCAGATCTGCTGCAAACTCTTCAGAACTGCCAAACAAACCCACTCACTGAGA	810			
QY	701	GGAAAGACGCCACTGGCGAGGCTGCTTACGCTGTGATGCTGTGGTGGCGCTGGTGG	760			
Db	811	AATCTGGTGAACAAAGGCTCTCAACACATTTATCTTATTATTGTTGTTGTTTC	870			
QY	761	TCTGCTTACGCCCTACCCACTCAACATCAAGCAGTTTCATG	801			
Db	871	TCTGTTTTCACACCTTACCATTGTTGCAATTTATCAACATATG	911			

RESULT 7
 CF359817/c
 LOCUS 636 bp mRNA linear EST 25-AUG-2003
 DEFINITION 821090 MARC 3P1G Sus scrofa cDNA 3', mRNA sequence.
 ACCESSION CF359817
 VERSION CF359817.1 GI:34156919
 KEYWORDS EST.
 SOURCE Sus. scrofa (pig)
 ORGANISM Sus. scrofa
 Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 636)
 Smith, T.P.L., Freking, B.A., Ford, J.J., Vallet, J.L., Fox, J.,
 Wise, T.A., Nonneman, D.J., Wray, J.E. and Keele, J.W.

TITLE A second set of porcine ESTs from a pooled-tissue normalized library
 JOURNAL Unpublished (2003)
 COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified with cross_match v0.990329.
 Plate: SRG8011 row: N column: 14
 Seq primer: TAGAAGCAGCTCCAGG.
 Location/Qualifiers
 1. .636
 /organism="Sus scrofa"
 /mol_type="mRNA"
 /db_xref="taxon:9823"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 3P1G"
 /note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI; Library made with RNA pooled from multiple tissues including brain, liver, muscle, placenta/endometrium, ovary, testes, and bone marrow."

FEATURES

source

ORIGIN

Query Match 14.4%; Score 159.4; DB 14; Length 636;
 Best Local Similarity 54.8%; Pred. No. 7.1e-23;
 Matches 316; Conservative 0; Mismatches 261; Indels 0; Gaps 0;

QY	101	CTCTGTTCTACAGACCCCTCTGCTTTCAGTGCCTCGGGAACATCTTGCCTTTGCC	160
Db	583	CTCTGATTTACAGATCTGCTTTCATATCGGCTCTCGGAACTTGTGGCCCTGATTG	524
QY	161	TTACCTGTCAAAAGACAGAGATCACTGCACAGGATCTACCTGGTGCACCTGGCTG	220
Db	523	TCATTTATCAAAACAGGAAAAATCACTCTACCACTATATTCACAACTTGGTGA	464
QY	221	TGCTGACCTGCTGTTTACCGTGGCTTACCGGAAGGTTGTGTATGTCTGGCT	280
Db	463	TTTCGGATATGATTTTACCACTGCTCTGCTACAGGATAGCTACTATGCTGGCT	404
QY	281	CCAGCTGGCTTTCCGCAAGGGCTCTGCAGGCTGACGGCTTTGTGCTCTACACGACA	340
Db	403	TTGACTGGAGATTGGCGAGGCTTGTGCAGGATACTGCTCTCGTGTATTATCAACA	344
QY	341	CCTACGGGGGGTCTACCTCATGCTGTGAGCGTGGACCATTTACCCAGCTGTGTCT	400
Db	343	CATACGAGCGGTGAATTCATGACCTGCTGAGTATGACCGTCTTGTGTGTGCTGC	284
QY	401	GTGCCACTGGGGCCCGGCTCTCCGACAGGCTGGCCCGCCAGGCTGGTCTGGTGGCCA	460
Db	283	ACCCCTGGGTACACAGATAAAAGATTGAACATGCAAAAGTGCATCTGCATATGCG	224
QY	461	TCTGACCTTGTGCTGCTGCACAGATGCCCTTGTGTGTCCTTGTGTCCTGATGCCA	520
Db	223	TCTGGATTCTTGATTTGCTCAACACTCCCGCTGCTCATAAACCCATCTCAAGCAGG	164
QY	521	TGTTGGCAAGCTGGCTGCTGATGAGTACAGACGATGAGTCACTCTCGGCTGCCCC	580
Db	163	AGGCTAAAGGACTACCTGCTGATGAGTATCCAACTTTGAGGAAACACAATCCCTCCCT	104
QY	581	TCATGGTCTGGTGGCTTTGCCATTTGGCTTCTGTGTGGCCAGTGGGATCATCTGTCT	640
Db	103	GGATTCTGCTGGGCGCTGTTTTCATAGGGTATGTAATTCGCTCTTAATCATCTCATCT	44
QY	641	GCTATATGAAGATCACTGGAAGCTGTGCAGCACAGC	677
Db	43	GCTACTCTCAGATCTGTGCAACTCTTCAAAAGCGC	7

```
RESULT 8
AL551903      1201 bp      mRNA      linear      EST 31-MAY-2003
LOCUS      AL551903 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION      clone CSOD1060X18 5-PRIME, mRNA sequence.
ACCESSION      AL551903
VERSION      AL551903.2 GI:31273719
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      1 (bases 1 to 1201)
JOURNAL      Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
COMMENT      Full-length cDNA libraries and normalization
              Unpublished (2001)
              On Feb 15, 2001 this sequence version replaced gi:12890291.
              Contact: Genoscope
              Genoscope - Centre National de Sequencage
              BP 191 91006 Evry cedex - France
              Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
              Library was constructed by Life Technologies, a division of
              Invitrogen. This sequence belongs to sequence cluster 6994.r For
              more information about this cluster, see
              http://www.genoscope.cns.fr/
              cgi-bin/cluster.cgi?seq=CSOD1060X18&cluster=6994.r. Contact :
              Feng Liang Email : fliang@lifetech.com URL :
              http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
              Faraday Avenue Genoscope sequence ID : CSOD1060DE09Qp1.
              Location/Qualifiers
                1. .1201
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="CSOD1060X18"
                  /tissue_type="PLACENTA COT 25-NORMALIZED"
                  /note="1st strand cDNA was primed with a NotI-oligo(dt)
                  primer. Five prime end enriched, double-strand cDNA was
                  digested with Not I and cloned into the Not I and EcoR V
                  sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES             source
   source
   Query Match      14.4%; Score 159.4; DB 9; Length 1201;
   Best Local Similarity 51.5%; Pred. No. 8.4e-23;
   Matches 361; Conservative 2; Mismatches 338; Indels 0; Gaps 0;

QY 101 CTCGTCTTACACAGCCCTCTGCTTCTCAGTGCCTGGGAAACATCTTGGCCCTTTGCC 160
DB 226 CTCTGCATTACAGCCTCTCTCATCTATGGCTGGTGGGAACTTACTAGCCTTGGTCG 285

QY 161 TTACTGTCAAAGAGCAGGAGATCACTGCACAGGCATCTACTGGTGACCTGGCTG 220
DB 286 TCATTGTTCAAACAGGAAAAAATCACTTACACCCCTCTATTCAACAAATTTGGTGA 345

QY 221 TGTCTGACCTGTGTTACCGTGGCTTACCGGAGGCTGGTGTGTATGCTGTGGCT 280
DB 346 TTCTTGATATCTTTTACACCGCTTTCCTTACAGNATAGCCTACTATGCAATGGCT 405

QY 281 CCAGCTGGCCCTTTCCGCAAGGGGCTCTGCAGGCTGACGGCTTTGTGCTCTACACGACA 340
DB 406 TTGACTGGAGATCGAGATGCTTGTGTAGATAACTGCGYAGTGTTTTACATCAACA 465

QY 341 CTTACGGGGGGTCTACTCTATGGCTGTGTAGCGTGGACCATACCCAGCTGTGCT 400
DB 466 CATATSCAGGTGTGAACCTTTATGACCTCGCTGAGTATGACCGCTTCTTGTGTTGTC 525

QY 401 GTGCCCACCTGGGGCGCGCCCTCCGACAGGCTGGCGCCGACGAGCTGTCTCGTGGCCA 460
DB 526 ACCCTCTACGCTCAACAGATATAAAGAGATTGAACATGCAAAAGGCGTGTGCATTTTG 585

QY 461 TGTGACCTTGGTGTGTGTGTGAGAGATGCCCTTCTTGTATGCCCATGACCAAGCCG 520
DB 585
```

```
Db 586 TCTGATTCTAGTATTGCTCAGACACTCCCACTCTCTCATCAACCCCTATGTCAAGCAGG 645
QY 521 TGTGGGCAACTGGCCTGATGAGTACAGCAGCATGGAGTCAGTCTCGGGCTGCC 580
DB 646 AGGTGAAAGATTACATGCAATGAGTATCCAACTTTGAAGAACTAAATCTCTCCT 705
QY 581 TCATGGTCTGCTGGCTTTGCCATTTGCTTCTGTGGCCAGTGGGATCAATCTGTCT 640
DB 706 GGAATCTGCTTGGGGCATGTTTCATAGGATGTACTTCCACTTATATATCTCTCATCT 765
QY 641 GCTATATGAATCACTGGAGCTGTGCACAGCTCGGAGAGCCAGCTGACCA 700
DB 766 GCTATCTCAGATCTGCTGCAACTCTTTCAGAACTGCAAAACAAACCACTACTGAGA 825
QY 701 GGAAGAGAGCCACTGGGAGCTGCCCTGCTTACGCTGCTGATGTGTGGCCCTGTGG 760
DB 826 RATCTGGTGTAAACAAAAGGCTCTCAACAATTTCTTATTATTGTTGTTGTTTC 885
QY 761 TCTGTTTCAAGCCCTTACCCTCAACATCAACAGAGTTTCATG 801
DB 886 TCTGTTTCAACACTTACCCTTGCATTTATTTTCAACATATG 926

RESULT 9
BM549799
LOCUS      BM549799
DEFINITION      AGENCOURT_6543842 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5745818
              5', mRNA sequence.
ACCESSION      BM549799
VERSION      BM549799.1 GI:18785503
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      1 (bases 1 to 1088)
JOURNAL      NIH-MGC http://mgi.nci.nih.gov/.
COMMENT      National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished (1999)
              Contact: Robert Strausberg, Ph.D.
              Email: cgsbbs-remail.nih.gov
              Tissue Procurement: Life Technologies, Inc.
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM12769 row: 9 column: 03
              High quality sequence stop: 619.

FEATURES             source
   source
   1..1088
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="IMAGE:5745818"
     /tissue_type="leukocyte"
     /lab_host="DH10B"
     /clone_lib="NIH_MGC_118"
     /note="Vector: PCMV-SPORT6; Site 1: NotI; Site 2: EcoRV
     (destroyed); RNA source leukocytes from anonymous pool of
     non-activated adult donors. Library is oligo-dT primed
     and directionally cloned (EcoRV site is destroyed upon
     cloning). Average insert size 1.7 kb, insert size range
     1.2-3.3 kb. Library is normalized and enriched for
     full-length clones and was constructed by C. Gruber
     (Invitrogen). Research Genetics tracking code 027. Note:
     this is a NIH_MGC Library."

ORIGIN
Query Match      14.1%; Score 155.6; DB 12; Length 1088;
Best Local Similarity 53.8%; Pred. No. 5.1e-22;
Matches 320; Conservative 0; Mismatches 275; Indels 0; Gaps 0;
```

101	QY	CTCTGTTCTACACAGCCCTCTCGTCTTCAGTGCCCCGGGAAACATCTCTTGCCCTTTGCC	160
130	Db	CTCTGCATTACAGCCTCGTCTTCATCATTTGGGCTCGTGGGAACTTACTAGCCTTTGGTGG	249
161	QY	TTACTGTCAAAAGACAGGAGAAATCAACTGCACAGGCATCTACCTGGTGCACCTGGGCTG	220
250	Db	TCATTTGTTCAAAAACAGGAAAAAATCAACTCTACCCCCCTCTATTCAACAAATTTGGTTA	309
221	QY	TGTCGACCTGCTGTTACCGTGGCCTTACCGGGAAGGGTGGTGTTATGTGCTGGGCT	280
310	Db	TTTCTGTATATACTTTTACCACCGCTTTGGCTACAGAAATAGCCTACTATGCAATGGGCT	369
281	QY	CCAGCTGCGCTTTCCGCAAGGGGCTCTGCAGGCTGACGGGTTTGCTGTACACCCGACA	340
370	Db	TTGACTGGAGAAATCGGAGATGCTTTGTGTAGGATTAATCGCGCTAGTGTTTTATCAACA	429
341	QY	CCTACGGGGGGGTCTA CCTCATGCGCTGTGTGAGCGTGGACCATTAACCCAGCTGGTCT	400
430	Db	CATATGCAGGTGTGAACCTTTATGACCTTCGCTGAGTATTGACCGCTTCATTGCTGTGGTGC	489
401	QY	GTGCCACTGGGGGCCCGCGCTCGCAGCGCTGGCGCGCCAGCGGTGCTCGGTGGCCA	460
490	Db	ACCCCTCTACGCTACAAACAGATAAAAGGATTTGAATGCAAAAGCGGTGTGCATTTTG	549
461	QY	TCTGGAACCTTGGTGCTGCTGCAGACGATGCCCTTGCTTTGATGCCCATGACCAAGCCGC	520
550	Db	TCTGGATTCTAGTATTTTGCTCAGACACTCCCACTCTCATCAACCTATGTCAAAGCAGG	609
521	QY	TGTTGGGCAAGCTGGCGCTGCATGAGATACAGACGATGAGTCAGTCTCGGGCTGCCCC	580
610	Db	AGGCTGAAGGATTACATGCATGGAGTATCCAAACTTTGAAGAAACTANATCTCTCCCT	669
581	QY	TCATGGTCCCTGGTGGCCCTTTGCGCATTTGGCTTTCTGTGGCGCAGTGGGGATCATCCTGTCT	640
670	Db	GGATTCTGCTTGGGGCATGTTTCATAGGATATGTACTTCCACTATATCATCTCATCT	729
641	QY	GCTATATGAAGATCACTGGAGAGCTGTGCAGCACAGCTCGGGAGGACCCAGTGAC	695
730	Db	GCTATTCTCAGATCTGTTGCAAACTTTTCAGAACTGCGAAAACAAAACCCACATCAC	784

RESULT 10
 BY765663
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS

BY765663 mRNA linear EST 17-DEC-2002
 CDNA full-length enriched, activated spleen Mus musculus
 clone P83003222 3', mRNA sequence.
 BY765663
 BY765663.1 GI:27203835
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 715)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaido, I., Osato, N., Saito, K., Suzuki, H., Yamanaka, I.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gobjori, T., Baldarelli, R., Hill, D.P., Bult, C.,
 Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
 Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
 Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
 Gariboldi, M., Gliss, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustavich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
 Kawaji, H., Kawasawa, Y., Kedierski, R.M., King, B.L., Konagaya, A.,
 Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G.,
 Petrovsky, N., Pilla, R., Pontius, J.U., Qi, D., Ramachandran, S.,
 Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
 Sandelin, A., Schneider, C., Sempie, C.A., Setou, M., Shimada, K.,
 Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,

RESULT 11

BY745873	LOCUS	BY745873	963 bp	linear	EST 17-DEC-2002
BY745873	DEFINITION	BY745873	RIKEN full-length enriched, 2 days neonate thymus thymic cells (NOB) Mus musculus cDNA clone E430001G12 5', mRNA sequence.	mRNA	thymus thymic

ACCESSION
BY745873
CELLS (NOV) MUS MUSCULUS C

VERSION
BY745873.1 GI:27173282

KEYWORDS EST.

SOURCE
Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chorda

Mammalia; Eutheria; Rodentia

AUTHORS
Okazaki, Y., Furuno, M., Kasai
Nikaido, T., Ogata N., Saito

Nikaido, I., Ushio, N., Saito
Kiyosawa H Yagi K Toma

Kiyosawa, H.; Yagi, K.; Ioma
schonbach, C.; Gojobori, T.

Hume, D.A.: Quackenbush, J.: Schombach, C.: Gelfond, I.: 1999

Batalov, S., Beisel, K.W., Brame, D.R., Gackowski, S.

Chothia, C.; Corbani, L.E.;

Fletcher, C.F., Forrest, A.,

Gariboldi, M., Gissi, C., Go

Gustíncich, S., Hirokawa, N.

Kawaji, H., Kawasaki, Y., Ke

Maltais, L., Marchionni, L.,

Numata, K., Okido, T.,
Pavan

Ravasi, T., Reed, J.C., Reed

Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Urtado, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kitahara, T., Konno, H., Nakamura, M., Sakazume, N., Satoh, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shiragawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683
12466851

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome.res@cg.riken.go.jp
URL: <http://genome.gsc.riken.go.jp/>
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,
Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,
Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawaji, J., Kojima, Y.,
Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murai, M., Nakamura, M.,
Nomura, K., Numasaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,
Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,
Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences *Mamm. Genome* 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes *Genome Res.* 10 (10), 1617-1630 (2000)

SIRKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer *Genome Res.* 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

These issues were provided by Dr. John Todd (Dept. of Medical Genetics, Wellcome Trust Centre for Molecular Mechanisms in Disease, Wellcome Trust/MRC building Addenbrookes Hospital, Cambridge) whose assistance we gratefully acknowledge. Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

```

Location/Qualifiers
1.. 963
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NOD"
/db_xref="taxon:10090"
/cldns="E430001G12"
/tissue_type="thymus"
/cell_type="thymic cells"
/clone_lib="RIKEN full-length enriched, 2 days neonate
thymus-thymic cells (NOD)"

```

Match 13.7%; Score 151.2; DB 13; Length 963;
 al Similarity 55.5%; Pred. No. 4.2e-21;
 331; Conservative 0; Mismatches 263; Indels 2; Gaps 2;

101 CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCTGGGAAACATCTTGCCCTTTGCC 160

```

170 CTCTGCAATCAGCCTGCTCTTCATCATTTGGGCTGGTGGAAACCTGCTGGGCTGGTTG 229
161 TTACTCTCTCAAAAGAGCAGGAAGATCAACTGACAGGCAATCACTGCTGGTGGCACTGGGCTG 220
230 TCATTGTTCAAAACAGAAAATAAATCAACTCAACCACTCTCTATTCAATGAATCTGCTCA 289
221 TGTCTGACCTGCTCTTACCGTGGCTTACCGGGAAGGTTGTTGTTATGCTGGGCT 280
290 TTTCTGACATCCTGTTTACCAACAGCTTTACCCACTCGGATAGCCTACTATGCTGGGCT 349
281 CCAGCTGGCCTTTTCGGCAAGGGCTCTGACAGGCTGACGGGCTTGTGCTCTACACCGACA 340
350 TTGATTGGAGATAGGTGATGCTCTGCGGGTAACTGCTCTGGTGTCTTACATCAACA 409
341 CCTACGGGGGGTCTACTCTATGCTGCTGACAGATGCTCTGCTGATGCGCATACCACTGCTG 400
410 CGTACGGAGGTGTAATCTTACGATTTGCTTGAGCATAGACCGCTTCTTTCGCTGGTGC 469
401 GTGCCCACTGGGGCCCGGCTCGCACGGCTCGCGCGCTGGCGGCTGCTGCTGGGCA 460
470 ACCTCTGGCTTACAAAGATT-AAAAGAAATCGAATAGCAAGGCTGCTGCTGCTGG 528
461 TCTGACCTTGGTCTGCTGACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 520
529 TCTGATCTGGCTTTGCTCAAACTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 588
521 TGTGGGCAAGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 580
589 AGGAGACAAGACCACTTGCATGAGTATCCAACTTGA-AGGACACGCTGCTGCTGCTG 647
581 TCATGCTGCTGGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 640
648 GGATCTGCTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 707
641 GCTATATGAGATCACCTGGAAGCTTGCAGACACCTCGGAGGACCCAGTGACC 696
708 GTTACTCTGAGTCTGCTGCAAACTCTCCAGGACTGCCAAGCAGAACCACTCACTCACC 763

```

```

RESULT 12
LOCUS BX427288 1201 bp mRNA linear EST 15-MAY-2003
DEFINITION BX427288 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
CSODM009YF17 5-PRIME, mRNA sequence.
ACCESSION BX427288
VERSION BX427288.1 GI:30786637
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1. (bases 1 to 1201)
L1, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope -- Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6994.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODM009CC09QPl&cluster=6994.r. Contact :
Feng Liang Email: fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODM009CC09QPl.
Location/Qualifiers
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

```

```

/clone="CSODM009YF17"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/notes="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with NotI and
cloned into the NotI and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
ORIGIN
Query Match 13.6%; Score 150.6; DB 13; Length 1201;
Best Local Similarity 53.6%; Pred. No. 5,9e-21;
Matches 309; Conservative 2; Mismatches 266; Indels 0; Gaps 0;
QY 101 CTCTGTTCTACAGACCCCTCTGCTCTTCTAGTCCCTGGGAAACATCCTTGGCCCTTTGCC 160
DB 268 CTCTGCAATTACAGCCCTGCTCTTCTCATTTGGGCTGCTGGGAACTTACTAGCTTGGTGC 327
QY 161 TTACCTGTCTCAAAAGAGCAGGAAGATCAACTGACAGGCAATCACTGCTGGTGCACCTGGTGTG 220
DB 328 TCATTGTTCAAAACAGGAAGAAATCAACTCTACACCTCTATTCAACAAATTTGGTGA 387
QY 221 TGTCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 280
DB 388 TTTCTGATTAATCTTTTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 447
QY 281 CCAGCTGGGCTTTTCGGCAAGGGCTCTGACAGGCTGACGGGCTTTGCTGCTCTACACCGACA 340
DB 448 TTGACTGGAGATCGAGATGCTTGTGTAGGATAAATCGCTAGTGTGTTTACATCAACA 507
QY 341 CCTACGGGGGGTCTACTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 400
DB 508 CATATGACAGGTGTAACCTTTATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 567
QY 401 GTGCCCACTGGGGCCCGGCTTCCGACGGCTCGCGCGCTGGCGGCTGGTGTGCTGCTGGGCA 460
DB 568 ACCCTCTACGCTTACAAAGATTAAGAGGATTTGAACATGCAAAAGGCGTGTGCTATATTG 627
QY 461 TCTGGACCTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 520
DB 628 TCTGGATCTTCTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 687
QY 521 TGTGGGGCAAGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 580
DB 688 AGGCTGAAAGGATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 747
QY 581 TCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 640
DB 748 GGATCTGCTTGGGCAAGTTCATGAGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 807
QY 641 GCTATATGAGATCACCTGGAAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 677
DB 808 GCTATTTCTGAGATCTGCTGCAAACTCTTTCAGAACTGC 844

```

```

RESULT 13
LOCUS AI659965/c 594 bp mRNA linear EST 18-DEC-1999
DEFINITION we64e01.x1 Soares thymus_NH5TH Homo sapiens cDNA clone
IMAGE:2345880 3', similar to SW:EBI2 HUMAN P32249 EBV-INDUCED G
PROTEIN-COUPLED RECEPTOR 2 ;, mRNA sequence.
ACCESSION AI659965
VERSION AI659965.1 GI:4763535
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

```


Search completed: February 21, 2004, 21:26:24
Job time : 2800 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 20:39:45 ; Search time 4338 Seconds
(without alignments)
11030.592 Million cell updates/sec

Title: US-10-041-615-33

Perfect score: 1104

Sequence: 1 atgataaactaggccctta.....gtgtcattgtttatgact 1104

Scoring table: **GAP60_NUC**
Gap60 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 10

Total number of hits satisfying chosen parameters: 2871005

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.cv.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1104	100.0	1104	6	AX702273	Sequence
2	461	41.8	4319	6	AX647101	AX647101 Sequence
C 3	461	41.8	98861	9	AL136961	AL136961 Human DNA
4	368	33.3	151174	2	AC016189	AC016189 Homo sapi
5	176	15.9	536	6	AX230148	AX230148 Sequence
C 6	171	15.5	151174	2	AC016189	AC016189 Homo sapi
C 7	159	14.4	556	6	AX244735	AX244735 Sequence
C 8	34	3.1	147559	10	AC117702	AC117702 Mus muscu
9	34	3.1	242813	2	AC095386	AC095386 Rattus no
10	33	3.0	227	8	AY023453	AY023453 Oryza sat
C 11	33	3.0	1695	8	AK067772	AK067772 Oryza sat
C 12	33	3.0	2790	8	BT010460	BT010460 Arabidops
13	33	3.0	59159	2	AC100332	AC100332 Mus muscu
14	33	3.0	78244	2	AC140499	AC140499 Mus muscu
15	33	3.0	98286	2	AP004394	AP004394 Oryza sat
16	33	3.0	116540	2	AC138355	AC138355 Mus muscu
17	33	3.0	136949	10	AL845271	AL845271 Mouse DNA
18	33	3.0	14422	8	AC007258	AC007258 Arabidops
19	33	3.0	146963	2	AC138591	AC138591 Mus muscu
C 20	33	3.0	150495	2	AC046148	AC046148 Mus muscu
21	33	3.0	152515	10	AL928608	AL928608 Mouse DNA
C 22	33	3.0	159988	2	AP005823	AP005823 Oryza sat
C 23	33	3.0	170979	2	AC026270	AC026270 Homo sapi
24	33	3.0	171030	9	CNS01DSF	AL121774 Human chr
25	33	3.0	175377	2	AC120929	AC120929 Rattus no
26	33	3.0	181037	2	AC116729	AC116729 Mus muscu
27	33	3.0	184867	2	AC109147	AC109147 Mus muscu
C 28	33	3.0	188382	2	AC118696	AC118696 Mus muscu
C 29	33	3.0	191781	10	AC124765	AC124765 Mus muscu
30	33	3.0	193437	2	AC119524	AC119524 Rattus no
31	33	3.0	203409	2	AC123277	AC123277 Rattus no
C 32	33	3.0	205468	2	AC141491	AC141491 Rattus no
33	33	3.0	210156	2	AC141495	AC141495 Rattus no
C 34	33	3.0	210988	2	AC139449	AC139449 Rattus no
35	33	3.0	227492	10	AC116585	AC116585 Mus muscu
36	33	3.0	228380	2	AC133352	AC133352 Rattus no
37	33	3.0	248276	2	AC103007	AC103007 Rattus no
38	33	3.0	252296	2	AC108538	AC108538 Rattus no
C 39	33	3.0	255956	10	AC110237	AC110237 Mus muscu
40	33	3.0	259141	2	AC117864	AC117864 Rattus no
41	33	3.0	283450	2	AC127733	AC127733 Rattus no
42	33	3.0	286145	2	AC099222	AC099222 Rattus no
43	33	3.0	292748	2	AC108574	AC108574 Rattus no
C 44	33	3.0	332036	2	AC129629	AC129629 Rattus no
45	32	2.9	59698	2	AC116398	AC116398 Mus muscu

ALIGNMENTS

RESULT 1
AX702273
LOCUS AX702273 1104 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 33 from Patent WO02064793.
ACCESSION AX702273
VERSION AX702273.1 GI:29537547
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Casman,S.J., Edinger,S.R., Ellerman,K., Smithson,G., Kekuda,R. and
Muralidhara,P.
TITLE Novel gpcr-like proteins and nucleic acids encoding same

[illegible]

```
Db 2940 CGCTCTCTGTTACACAGCCCTCTCTGCTTTCAGTGCCCTGGAAACATCCTTGCCCTT 2999
Qy 157 TGCCTTACCTGTCAAAGAGCAGGAAGATCACTGACAGGATCTACTCTGTGACCTG 216
Db 3000 TGCCTTACCTGTCAAAGAGCAGGAAGATCACTGACAGGATCTACTCTGTGACCTG 3059
Qy 217 GCTGTGCTGTGACCTGTCTTACCTGCGCTTACCGGGAAGGCTGTGTGTTATGTGCTG 276
Db 3060 GCTGTGCTGTGACCTGTCTTACCTGCGCTTACCGGGAAGGCTGTGTGTTATGTGCTG 3119
Qy 277 GGTCACAGTGGCTTTTCGGCAAGGGGCTCTGACGCTGACGGCTTTGTGCTTACACC 336
Db 3120 GGCTCCAGCTGGCTTTTCGGCAAGGGGCTCTGACGCTGACGGCTTTGTGCTTACACC 3179
Qy 337 GACACCTACGGGGGGGTCTACCTCATGCTGTGTGAGCGTGAACCATACCCAGCTGTG 396
Db 3180 GACACCTAC-GGGGGGTCTACCTCATGCTGTGTGAGCGTGAACCATACCCAGCTGTG 3238
Qy 397 GTCTGTGCCACTTGGGGGCCCGCCCTCCGACCGCTGCGCGCCAGGCTGTGTGCGTG 456
Db 3239 GTCTGTGCCACTTGGGGGCCCGCCCTCCGACCGCTGCGCGCCAGGCTGTGTGCGTG 3298
Qy 457 GCCATCTGACCTGTGTGCTGTGCGAGATCCCTTGTCTTGTATGCCCATGACCAAG 516
Db 3299 GCCATCTGACCTGTGTGCTGTGCGAGATCCCTTGTCTTGTATGCCCATGACCAAG 3358
Qy 517 CCGCTGTGTGGCAAGCTGGCTCTCATGTGATGACAGCAGCATGAGTCAGTCTCGGGCTG 576
Db 3359 CCGCTGTGTGGCAAGCTGGCTCTCATGTGATGACAGCAGCATGAGTCAGTCTCGGGCTG 3418
Qy 577 CCGCTGTGTGTGGCTGTGGCTTGGCATGGCTTCTGTGGCCAGTGGGATCATCCG 636
Db 3419 CCGCTGTGTGTGGCTGTGGCTTGGCATGGCTTCTGTGGCCAGTGGGATCATCCG 3478
Qy 637 TCCTGCTATATGAAGATCACTCGAAGCTGTGACGACACAGCT 678
Db 3479 TCCTGCTATATGAAGATCACTCGAAGCTGTGACGACACAGCT 3520

RESULT 3
AL136961/c
LOCUS Human DNA sequence from clone Rp11-178C10 on chromosome 13,
DEFINITION complete sequence.
ACCESSION AL136961
VERSION AL136961.19 GI:13620283
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 98861)
REFERENCE
AUTHORS Blakey,S.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
COMMENT
On Apr 12, 2001 this sequence version replaced gi:13276997.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
```

```
SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
Rp11-178C10 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOS: PBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
Rp11-178C10 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone Rp11-178C10 is at 98861 in this
sequence. The true right end of clone Rp11-461N23 is at 100 in this
sequence.
FEATURES
Location/Qualifiers
source
1..98861
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/clone="Rp11-178C10"
/clone_lib="RPCI-11.1"
repeat_region
161..463
/note="L1MB8 repeat: matches 5836..6160 of consensus"
repeat_region
656..723
/note="L2 repeat: matches 2592..2659 of consensus"
repeat_region
2385..2420
/note="WIR repeat: matches 111..146 of consensus"
repeat_region
4555..4646
/note="L2 repeat: matches 1031..1120 of consensus"
repeat_region
4989..5306
/note="MER7A repeat: matches 2..345 of consensus"
repeat_region
5859..5987
/note="L2 repeat: matches 2574..2708 of consensus"
repeat_region
6215..6548
/note="Charlie1 repeat: matches 2292..2629 of consensus"
repeat_region
6557..6635
/note="L2 repeat: matches 2128..2206 of consensus"
repeat_region
6730..7233
/note="L2 repeat: matches 1330..1889 of consensus"
repeat_region
8301..8499
/note="WIR repeat: matches 17..219 of consensus"
repeat_region
8893..8994
/note="L2 repeat: matches 1551..1656 of consensus"
repeat_region
9079..9126
/note="24 copies 2 mer at 75% conserved"
repeat_region
9554..9754
/note="MER20 repeat: matches 33..218 of consensus"
repeat_region
9816..9979
/note="FAM repeat: matches 2..165 of consensus"
repeat_region
10402..10703
/note="L1M1 repeat: matches 5264..5577 of consensus"
repeat_region
12181..12626
/note="L1 repeat: matches 3474..3949 of consensus"
repeat_region
12654..13012
/note="L1ME3 repeat: matches 5571..5927 of consensus"
repeat_region
13491..14076
/note="L1 repeat: matches 2226..2881 of consensus"
repeat_region
16310..17185
/note="L1MC4 repeat: matches 6237..7151 of consensus"
repeat_region
17207..17622
/note="L1MC4 repeat: matches 2..413 of consensus"
repeat_region
17628..18380
/note="MER47B repeat: matches 7197..7976 of consensus"
repeat_region
20797..21100
/note="AluY repeat: matches 1..297 of consensus"
repeat_region
22746..22814
/note="L2 repeat: matches 2574..2643 of consensus"
repeat_region
23758..23981
/note="AluDb repeat: matches 85..310 of consensus"
```

```
repeat_region 25108..25212
/note="I2 repeat: matches 2570..2688 of consensus"
repeat_region 25740..25795
/note="I2R repeat: matches 196..251 of consensus"
repeat_region 25982..26012
/note="I2R repeat: matches 2701..2731 of consensus"
repeat_region 26005..26208
/note="I2R repeat: matches 1772..1970 of consensus"
repeat_region 26209..26513
/note="AluSx repeat: matches 1..305 of consensus"
repeat_region 26514..28201
/note="I2R repeat: matches 1..1772 of consensus"
repeat_region 32616..32843
/note="6 copies 38 mer 86% conserved"
repeat_region 33333..35362
/note="15 copies 2 mer ac 100% conserved"
repeat_region 36220..36303
/note="2 copies 42 mer 97% conserved"
repeat_region 36967..37274
/note="AluSg repeat: matches 1..313 of consensus"
repeat_region 39481..39767
/note="AluSx repeat: matches 1..297 of consensus"
repeat_region 41084..41111
/note="14 copies 2 mer aa 89% conserved"
repeat_region 41119..41174
/note="I2R repeat: matches 1..58 of consensus"
repeat_region 41175..41482
/note="AluSg repeat: matches 1..309 of consensus"
repeat_region 41483..41785
/note="I2R repeat: matches 58..344 of consensus"
repeat_region 41791..41842
/note="I2 repeat: matches 1215..1266 of consensus"
repeat_region 42051..42573
/note="I2R repeat: matches 2..555 of consensus"
repeat_region 42857..43014
/note="I2 repeat: matches 2553..2733 of consensus"
repeat_region 43250..43279
/note="15 copies 2 mer ca 100% conserved"
repeat_region 44262..44708
/note="I2R repeat: matches 1..466 of consensus"
repeat_region 44860..45175
/note="I2R repeat: matches 1..328 of consensus"
repeat_region 45204..45235
/note="16 copies 2 mer tg 87% conserved"
repeat_region 45389..45559
/note="I2R repeat: matches 4..175 of consensus"
repeat_region 45639..45798
/note="I2R repeat: matches 9..184 of consensus"
repeat_region 45963..46162
/note="I2R repeat: matches 25..250 of consensus"
repeat_region 46281..46837
/note="I2R repeat: matches 1..510 of consensus"
repeat_region 47886..51275
/note="I2R repeat: matches 2744..6141 of consensus"
repeat_region 51364..51523
/note="I2R repeat: matches 1..189 of consensus"
repeat_region 51736..51791
/note="I2R repeat: matches 118..178 of consensus"
repeat_region 51848..52167
/note="I2R repeat: matches 86..422 of consensus"
repeat_region 53138..53915
/note="I2R repeat: matches 41..789 of consensus"
repeat_region 54321..54629
/note="AluSg repeat: matches 1..310 of consensus"
repeat_region 55247..55308
/note="I2R repeat: matches 24..87 of consensus"
repeat_region 55309..55622
/note="AluSg repeat: matches 1..313 of consensus"
repeat_region 55623..55715
/note="I2R repeat: matches 87..176 of consensus"
```

```
repeat_region 57171..57271
/note="I2R repeat: matches 437..541 of consensus"
repeat_region 57462..57560
/note="I2R repeat: matches 169..264 of consensus"
repeat_region 57561..58049
/note="I2R repeat: matches 1..519 of consensus"
repeat_region 58063..58287
/note="AluSg/x repeat: matches 163..300 of consensus"
repeat_region 58336..58475
/note="I2R repeat: matches 76..205 of consensus"
repeat_region 58496..58809
/note="AluSx repeat: matches 1..311 of consensus"
repeat_region 58818..59182
/note="I2R repeat: matches 5789..6157 of consensus"
repeat_region 59198..59336
/note="I2R repeat: matches 1..139 of consensus"
repeat_region 59337..59435
/note="I2R repeat: matches 169..264 of consensus"
repeat_region 59626..59726
/note="I2R repeat: matches 437..541 of consensus"
repeat_region 59804..60249
/note="I2R repeat: matches 5301..5782 of consensus"

Query Match 41.8%; Score 461; DB 9; Length 98861;
Best Local Similarity 99.8%; Pred. No. 4.1e-258;
Matches 581; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 97 CGCTCTGTTCTACACAGCCCTCTGCTCTCAGTGCCTCGGAAACATCTTGGCCTT 156
DB 56922 CGCTCTGTTCTACACAGCCCTCTGCTCTCAGTGCCTCGGAAACATCTTGGCCTT 56863

QY 157 TGCCTTACCTGTCAGAGAGAGAGATCAATGTCACAGGATCTACCTGGTGCACCTG 216
DB 56862 TGCCTTACCTGTCAGAGAGAGAGATCAATGTCACAGGATCTACCTGGTGCACCTG 56803

QY 217 GCTGTGTGACCTGCTTTCACCGTGGCTTACCGGAGGCTGGTGTATGTGCTG 276
DB 56802 GCTGTGTGACCTGCTTTCACCGTGGCTTACCGGAGGCTGGTGTATGTGCTG 56743

QY 277 GCTTCCAGTGGCCTTTCGGGAGGCTGTCAGAGTGCAGGCTTGTGCTCTACAC 336
DB 56742 GCTTCCAGTGGCCTTTCGGGAGGCTGTCAGAGTGCAGGCTTGTGCTCTACAC 56683

QY 337 GACACCTACGGGGGGCTTACCTCATGGCTGTGTGAGCGTGCACCATACCAAGTGTG 396
DB 56682 GACACCTAC-GGGGGGGCTTACCTCATGGCTGTGTGAGCGTGCACCATACCAAGTGTG 56624

QY 397 GTCTGTGCCACTGGGGGGCTTCCGACGCTGCGCGGCTGGCGCCGAGGCTGTGCTG 456
DB 56623 GTCTGTGCCACTGGGGGGCTTCCGACGCTGCGCGGCTGGCGCCGAGGCTGTGCTG 56564

QY 457 GCGATGTGACCTTGGTGTGCTGTCAGACGATGCTTGTGCTTGTGATGCCATGACCAAG 516
DB 56563 GCGATGTGACCTTGGTGTGCTGTCAGACGATGCTTGTGCTTGTGATGCCATGACCAAG 56504

QY 517 CGCTGTGGGCAAGCTGGGCTGCATGGAGTACAGAGATGAGTGCAGTCTCGGGCTG 576
DB 56503 CGCTGTGGGCAAGCTGGGCTGCATGGAGTACAGAGATGAGTGCAGTCTCGGGCTG 56444

QY 577 CCGCTCATGTGCTGGGCTTGGCTTGCATGGCTTGTGCGGCGAGTGGGATCATCTG 636
DB 56443 CCGCTCATGTGCTGGGCTTGGCTTGCATGGCTTGTGCGGCGAGTGGGATCATCTG 56384

QY 637 TCCTGTATATGAAGATCACCTGGAAGCTGTGAGACAGCT 678
DB 56383 TCCTGTATATGAAGATCACCTGGAAGCTGTGAGACAGCT 56342

RESULT 4
AC016189
LOCUS
DEFINITION Homo sapiens clone RP11-23M9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC016189
```

VERSION AC016189.3 GI-9129134
KEYWORDS HTG; HTGS; PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE 1 (bases 1 to 151174)
JOURNAL 2 (bases 1 to 151174)
REFERENCE 1 (bases 1 to 151174)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Becker, R., Boguslavsky, L., Bouckghalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gargyala, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McSwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thoman, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6939915.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997).
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4399
Center clone name: 23_M_9

* NOTE: This record contains 152 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 851: contig of 851 bp in length
* 852 951: gap of 100 bp
* 952 1804: contig of 853 bp in length
* 1805 1904: gap of 100 bp
* 1905 2719: contig of 815 bp in length
* 2720 2819: gap of 100 bp
* 2820 3668: contig of 849 bp in length
* 3669 3768: gap of 100 bp
* 3769 4657: contig of 889 bp in length
* 4658 4758: gap of 100 bp
* 4759 5600: contig of 843 bp in length
* 5601 5700: gap of 100 bp
* 5701 6532: contig of 832 bp in length
* 6533 6632: gap of 100 bp
* 6633 7481: contig of 849 bp in length
* 7482 7581: gap of 100 bp
* 7582 8447: contig of 866 bp in length
* 8448 8547: gap of 100 bp
* 8548 9383: contig of 836 bp in length
* 9384 9484 10342: contig of 859 bp in length
* 10343 10442: gap of 100 bp
* 10443 11233: contig of 881 bp in length
* 11234 11423: gap of 100 bp
* 11424 12254: contig of 831 bp in length
* 12255 12354: gap of 100 bp
* 12355 13288: contig of 934 bp in length
* 13289 13388: gap of 100 bp
* 13389 14235: contig of 847 bp in length
* 14236 14335: gap of 100 bp
* 14336 15172: contig of 837 bp in length
* 15173 15272: gap of 100 bp
* 15273 16130: contig of 858 bp in length
* 16131 16230: gap of 100 bp
* 16231 17073: contig of 843 bp in length
* 17074 17173: gap of 100 bp
* 17174 18003: contig of 830 bp in length
* 18004 18103: gap of 100 bp
* 18104 18939: contig of 836 bp in length
* 18940 19039: gap of 100 bp
* 19040 19885: contig of 846 bp in length
* 19886 20934: contig of 849 bp in length
* 20935 21761: contig of 827 bp in length
* 21762 21861: gap of 100 bp
* 21862 22743: contig of 882 bp in length
* 22744 22843: gap of 100 bp
* 22844 23709: contig of 865 bp in length
* 23709 24676: contig of 868 bp in length
* 24677 24776: gap of 100 bp
* 24777 25602: contig of 826 bp in length
* 25603 25702: gap of 100 bp
* 25703 26560: contig of 858 bp in length
* 26561 26660: gap of 100 bp
* 26661 27504: contig of 844 bp in length
* 27505 28449: contig of 845 bp in length
* 28450 28549: gap of 100 bp
* 28550 29398: contig of 849 bp in length
* 29399 29498: gap of 100 bp
* 29499 30357: contig of 859 bp in length
* 30358 30457: gap of 100 bp
* 30458 31323: contig of 866 bp in length
* 31324 31423: gap of 100 bp
* 31424 32300: contig of 877 bp in length
* 32301 32400: gap of 100 bp
* 32401 33254: contig of 854 bp in length
* 33255 33354: gap of 100 bp
* 33355 34218: contig of 864 bp in length
* 34219 34318: gap of 100 bp
* 34319 35169: contig of 851 bp in length
* 35170 35269: gap of 100 bp
* 35270 36125: contig of 856 bp in length
* 36126 36225: gap of 100 bp
* 36226 37061: contig of 836 bp in length
* 37062 37161: gap of 100 bp
* 37162 38003: contig of 842 bp in length
* 38004 38103: gap of 100 bp
* 38104 38945: contig of 842 bp in length
* 38946 39045: gap of 100 bp
* 39046 39882: contig of 837 bp in length
* 39883 39982: gap of 100 bp
* 39983 40933: contig of 851 bp in length
* 40934 41779: gap of 100 bp
* 41780 41879: contig of 846 bp in length
* 41880 42723: contig of 843 bp in length
* 42724 42823: gap of 100 bp
* 42823 43679: contig of 856 bp in length
* 43679 44634: gap of 100 bp
* 44635 44734: contig of 856 bp in length
* 44734 44635: gap of 100 bp


```

REFERENCE
AUTHORS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL      1 (Bases 1 to 151174)
REFERENCE    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
AUTHORS      Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
TITLE        Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
JOURNAL      Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
REFERENCE    Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Horton,L.,
AUTHORS      Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
TITLE        Heland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
JOURNAL      Lebowitz,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
REFERENCE    McSwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
AUTHORS      Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
TITLE        Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
JOURNAL      Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
REFERENCE    Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
AUTHORS      Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE        Direct Submission
JOURNAL      Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
COMMENT      Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6939915.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4399
Center clone name: 23_M_9
-----
* NOTE: This record contains 152 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 851: contig of 851 bp in length
* 852 951: gap of 100 bp
* 952 1804: contig of 853 bp in length
* 1805 1904: gap of 100 bp
* 1905 2719: contig of 815 bp in length
* 2720 2819: gap of 100 bp
* 2820 3668: contig of 849 bp in length
* 3669 3768: gap of 100 bp
* 3769 4657: contig of 889 bp in length
* 4658 4757: gap of 100 bp
* 4758 5600: contig of 843 bp in length
* 5601 5700: gap of 100 bp
* 5701 6532: contig of 832 bp in length
* 6533 6632: gap of 100 bp
* 6633 7481: contig of 849 bp in length
* 7482 7581: gap of 100 bp
* 7582 8447: contig of 866 bp in length
* 8448 8547: gap of 100 bp
* 8548 9383: contig of 836 bp in length
* 9384 9483: gap of 100 bp
* 9484 10342: contig of 859 bp in length
* 10343 10442: gap of 100 bp
* 10443 11323: contig of 881 bp in length
* 11324 11423: gap of 100 bp
* 11424 12254: contig of 831 bp in length

```

```

12255 12354: gap of 100 bp
12355 13288: contig of 934 bp in length
13289 13388: gap of 100 bp
13389 14235: contig of 847 bp in length
14236 14335: gap of 100 bp
14336 15172: contig of 837 bp in length
15173 15272: gap of 100 bp
15273 16130: contig of 858 bp in length
16131 16230: gap of 100 bp
16231 17073: contig of 843 bp in length
17074 17173: gap of 100 bp
17174 18003: contig of 830 bp in length
18004 18103: gap of 100 bp
18104 18399: contig of 836 bp in length
18400 19039: gap of 100 bp
19040 19885: contig of 846 bp in length
19886 20834: contig of 849 bp in length
20835 21761: contig of 827 bp in length
21762 21861: gap of 100 bp
21862 22743: contig of 882 bp in length
22744 22843: gap of 100 bp
22844 23709: contig of 865 bp in length
23710 23808: gap of 100 bp
23809 24676: contig of 868 bp in length
24677 25602: contig of 826 bp in length
25603 25702: gap of 100 bp
25703 26560: contig of 858 bp in length
26561 26660: gap of 100 bp
26661 27504: contig of 844 bp in length
27505 27604: gap of 100 bp
27606 28449: contig of 845 bp in length
28450 28549: gap of 100 bp
28550 29398: contig of 849 bp in length
29399 29498: gap of 100 bp
29499 30357: contig of 859 bp in length
30358 30457: gap of 100 bp
30458 31323: contig of 866 bp in length
31324 31423: gap of 100 bp
31424 32300: contig of 877 bp in length
32301 3254: contig of 854 bp in length
3255 33354: gap of 100 bp
33355 34218: contig of 864 bp in length
34219 35169: contig of 851 bp in length
35170 35269: gap of 100 bp
35270 36125: contig of 856 bp in length
36126 36225: gap of 100 bp
36226 37061: contig of 836 bp in length
37062 37161: gap of 100 bp
37162 38003: contig of 842 bp in length
38004 38103: gap of 100 bp
38104 38945: contig of 842 bp in length
38946 39045: gap of 100 bp
39046 39882: contig of 837 bp in length
39883 39982: gap of 100 bp
39983 40833: contig of 851 bp in length
40834 40934: gap of 100 bp
40934 41779: contig of 846 bp in length
41780 41879: gap of 100 bp
41880 42722: contig of 843 bp in length
42723 42823: gap of 100 bp
42823 43679: contig of 856 bp in length
43679 44634: gap of 100 bp
44635 44734: contig of 856 bp in length
44735 45587: gap of 100 bp
45588 45688: gap of 100 bp
45689 4642: contig of 855 bp in length
4642 46543: gap of 100 bp

```

[illegible]

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;
1

REFERENCE
AUTHORS Vogeli, G. and Wood, L. S.
TITLE G protein-coupled receptors
JOURNAL Patent: WO 016750-A 64 13-SEP-2001;
PHARMACIA & UPJOHN COMPANY (US)

FEATURES
source Location/Qualifiers
1. 556
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 14.4%; Score 159; DB 6; Length 556;
Best Local Similarity 99.5%; Pred. No. 6e-81;
Matches 209; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 697 AGCAGGAAGAGGAGCCACCTGGGAGGCTGCTGCTTACGCTGCTGATGCTGTCGTCGTCG 756
|||
Db 556 AGCAGGAAGAGGAGCCACCTGGGAGGCTGCTGCTTACGCTGCTGATGCTGTCGTCGTCG 497
|||
QY 757 GTGCTGCTGCTTACGAGCCCTACCACTTCAACATCAAGCAGTTTCATGGCGAGAGGATGCTC 816
|||
Db 496 GTGCTGCTGCTTACGAGCCCTACCACTTCAACATCAAGCAGTTTCATGGCGAGAGGATGCTC 437
|||
QY 817 CACCTGCCATCTGTGCGGAGGAGGGCTTCTTACTGTCTGCTTACGCCACACGTCGCC 876
|||
Db 436 CACCTGCCATCTGTGCGGAGGAGGGCTTCTTACTGTCTGCTTACGCCACACGTCGCC 377
|||
QY 877 CTCATGAACATGAAGTGTGGCATTAACCCCA 906
|||
Db 376 CTCATGAACATGAAGTGTGGCATTAACCCCA 347
|||

RESULT 8
AC117702/c
LOCUS
DEFINITION Mus musculus chromosome 10, clone RP23-46412, complete sequence.
ACCESSION AC117702
VERSION 1 (bases 1 to 147559)
KEYWORDS HTG.
SOURCE Mus musculus (house mouse).
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 147559)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus chromosome 10, clone RP23-46412
Unpublished
2 (bases 1 to 147559)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguski, L.,
Bouckhalter, B., Brown, A., Camarata, J., Campolongo, A., Chang, J.,
Chazaro, B., Choelp, Y., Collangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, S., Dodge, S.,
Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gored, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Illiev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R.,
Landers, T., Lebeckzy, L., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, C.,
Mihova, T., Mlenga, P., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,

Only the last 147.6 kilobases of this clone are being submitted. The remainder overlaps accession number AC113207 [WICGR project L23741].

Location/Qualifiers
1..147559 /organism="Mus musculus" /mol type="genomic DNA"

* 1 237255: contig of 237255 bp in length


```

RESULT 13
AC100332
LOCUS      AC100332
DEFINITION Mus musculus clone RP23-126G16, DNA linear HTG 22-NOV-2001
ACCESSION AC100332
VERSION   AC100332.1 GI:17047698
KEYWORDS  HTG; HTGS PHASE0.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 59159)
AUTHORS   Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE     Mus musculus, clone RP23-126G16
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 59159)
AUTHORS   Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
           Anderson, S., Barna, N., Bastien, V., Boguski, L., Bouck, G.,
           Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
           Cho, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,
           Cooke, P., DeArillano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
           Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
           Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
           Hages, B., Heaford, A., Horton, L., Hulme, W., Iliiev, I., Johnson, R.,
           Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
           Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
           MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
           McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
           Meneub, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
           Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
           Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
           Raymond, C., Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P.,
           Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
           Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
           Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
           Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
           Viell, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
           Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
           Direct Submission
           Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           ----- Genome Center
           Center: Whitehead Institute/ MIT Center for Genome Research
           Center code: WTRB
           Web site: http://www-seq.wi.mit.edu
           Contact: sequence.submissions@genome.wi.mit.edu
           ----- Project Information
           Center project name: Li5119
           Center clone name: 126_G_16
           -----
           * NOTE: This record contains 74 individual
           * sequencing reads that have not been assembled into
           * contigs. Runs of N are used to separate the reads
           * and the order in which they appear is completely for
           * arbitrary. Low-pass sequence sampling is useful for
           * identifying clones that may be gene-rich and allows
           * overlap relationships among clones to be deduced.
           * However, it should not be assumed that this clone
           * will be sequenced to completion. In the event that
           * the record is updated, the accession number will
           * be preserved.
           * 1 704: contig of 704 bp in length
           * 705 804: gap of 100 bp
           * 805 1534: contig of 730 bp in length
           * 1535 1634: gap of 100 bp
           * 1635 2338: contig of 704 bp in length
           * 2339 2438: gap of 100 bp
           * 2439 3120: contig of 682 bp in length
           * 3121 3220: gap of 100 bp
           * 3221 3907: contig of 687 bp in length
           * 3908 4007: gap of 100 bp
           * 4008 4691: contig of 684 bp in length
           * 4692 4791: gap of 100 bp
           * 4792 5475: contig of 684 bp in length
           * 5476 5575: gap of 100 bp
           * 5576 6258: contig of 683 bp in length
           * 6259 6358: gap of 100 bp
           * 6359 7057: contig of 699 bp in length
           * 7058 7157: gap of 100 bp
           * 7158 7867: contig of 710 bp in length
           * 7868 7967: gap of 100 bp
           * 7968 8647: contig of 680 bp in length
           * 8648 9478: gap of 100 bp
           * 9479 9578: contig of 731 bp in length
           * 9579 10292: contig of 714 bp in length
           * 10293 10392: gap of 100 bp
           * 10393 11088: contig of 696 bp in length
           * 11089 11188: gap of 100 bp
           * 11189 11870: contig of 682 bp in length
           * 11871 11970: gap of 100 bp
           * 11971 12674: contig of 704 bp in length
           * 12675 12774: gap of 100 bp
           * 12775 13485: contig of 711 bp in length
           * 13486 13585: gap of 100 bp
           * 13586 14306: contig of 721 bp in length
           * 14307 14406: gap of 100 bp
           * 14407 15121: contig of 715 bp in length
           * 15122 15221: gap of 100 bp
           * 15222 15953: contig of 732 bp in length
           * 15954 16674: gap of 100 bp
           * 16674 16773: contig of 620 bp in length
           * 16774 17453: contig of 680 bp in length
           * 17454 17553: gap of 100 bp
           * 17554 18238: contig of 685 bp in length
           * 18239 18339: gap of 100 bp
           * 18339 19025: contig of 687 bp in length
           * 19026 19125: gap of 100 bp
           * 19126 19805: contig of 680 bp in length
           * 19806 19905: gap of 100 bp
           * 19906 20604: contig of 699 bp in length
           * 20605 20704: gap of 100 bp
           * 20705 21415: contig of 711 bp in length
           * 21416 21515: gap of 100 bp
           * 21516 22196: contig of 681 bp in length
           * 22197 22296: gap of 100 bp
           * 22297 23023: contig of 727 bp in length
           * 23024 23123: gap of 100 bp
           * 23124 23834: contig of 711 bp in length
           * 23835 23934: gap of 100 bp
           * 23935 24642: contig of 708 bp in length
           * 24643 24742: gap of 100 bp
           * 24743 25428: contig of 686 bp in length
           * 25429 25528: gap of 100 bp
           * 25529 26219: contig of 691 bp in length
           * 26220 26319: gap of 100 bp
           * 26320 27005: contig of 686 bp in length
           * 27006 27105: gap of 100 bp
           * 27106 27813: contig of 708 bp in length
           * 27814 27913: gap of 100 bp
           * 27914 28625: contig of 712 bp in length
           * 28626 28725: gap of 100 bp
           * 28726 29448: contig of 723 bp in length
           * 29449 29548: gap of 100 bp
           * 29549 30271: contig of 723 bp in length
           * 30272 30371: gap of 100 bp
           * 30372 31100: contig of 729 bp in length
           * 31101 31200: gap of 100 bp
           * 31201 31933: contig of 733 bp in length
           * 31934 32033: gap of 100 bp
           * 32034 32721: contig of 688 bp in length
           * 32722 32821: gap of 100 bp

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 19:27:14 ; Search time 483 Seconds
(without alignments)
9710.167 Million cell updates/sec

Title: US-10-041-615-33
Perfect score: 1104
Sequence: 1 atgataaaactaggccctta.....gtgcattgtctttatgact 1104

Scoring table: OLIGO_NUC
Gapop_60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 10

Total number of hits satisfying chosen parameters: 1958023

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N Geneseq_29Jan04.*
1: geneseqn1980s.*
2: geneseqn1990s.*
3: geneseqn2000s.*
4: geneseqn2001as.*
5: geneseqn2001bs.*
6: geneseqn2002as.*
7: geneseqn2003as.*
8: geneseqn2003bs.*
9: geneseqn2003cs.*
10: geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1104	100.0	1104	6	ABE64723 CDNA enco
2	461	41.8	4319	9	ADC86840 Human GPC
3	461	41.8	9409	4	AAK76477 Human imm
4	458	41.5	1001	6	AAK98131 Human DNA
5	409	37.0	1526	6	ABK94922 Human nov
6	318	28.8	330	6	AAK98111 Human DNA
7	279	25.3	1771	5	AAK78869 DNA enco
8	176	15.9	536	4	AAK42840 Human G P
9	159	14.4	556	4	AAK30791 Human CDN
10	103	9.3	594	6	AAK98061 Human DNA
11	33	3.0	22760	5	ABK15489 Human ner
12	31	2.8	232	4	AAI22451 Probe #12
13	31	2.8	232	4	ABK67527 Human foe
14	31	2.8	232	4	AAI47743 Probe #16
15	31	2.8	232	4	ABA49623 Human bre
16	31	2.8	232	4	ABA34608 Probe #13
17	31	2.8	232	4	AAK41698 Human bon
18	31	2.8	232	4	AAK15953 Human bra
19	31	2.8	232	4	ABK1287 Human liv
20	31	2.8	232	5	AAI08137 Probe #81
21	31	2.8	232	6	ABK15705 Human gen
22	31	2.8	384	4	AAI13216 Probe #31
23	31	2.8	384	4	ABA54915 Human foe

ALIGNMENTS

RESULT 1
ABE64723
ID ABE64723 standard; CDNA; 1104 BP.

XX AC ABE64723;

XX DT 15-NOV-2002 (first entry)

XX DE CDNA encoding human GPCR16 protein.

XX KW G protein coupled receptor related protein; ss: gene; human; GPCR;
KW cardiomyopathy; atherosclerosis; diabetes; cancer; stroke; sclerosis;
KW Von Hippel-Lindau syndrome; Alzheimer's disease; tuberculous sclerosis;
KW hypercalcaemia; Parkinson's disease; Huntington's disease;
KW cerebral palsy; epilepsy; Lesch-Nyhan syndrome; multiple sclerosis;
KW ataxia-telangiectasia; leukodystrophy; addiction; anxiety; depression;
KW pain; obesity; Crohn's disease; osteoporosis; haemophilia; asthma;
KW inflammatory bowel disease; infertility; hypertension; scleroderma;
KW arthritis; human immunodeficiency virus; autoimmune disease; HIV;
KW infection; graft-versus-host disease.

XX OS Homo sapiens.

XX FN WO200264793-A2.

XX PD 22-AUG-2002.

XX PF 03-JAN-2002; 2002WO-US0000056.

XX PR 03-JAN-2001; 2001US-0259552P.

XX PR 03-JAN-2001; 2001US-0260544P.

XX PR 20-MAR-2001; 2001US-0277405P.

XX PA (CURA-) CURAGEN CORP.

XX PI Casman SJ, Edinger SR, Ellerman K, Smithson G, Kekuda R;

XX PI Muraidhara P;

XX DR WPI; 2002-643487/69.

XX DR P-PSDB; ABG32286.

XX PT New isolated G protein coupled receptor polypeptides and polynucleotides,
XX PT useful in gene therapy, particularly for treating or preventing
XX PT cardiomyopathy, atherosclerosis, diabetes, multiple sclerosis, Crohn's
XX PT disease or cancer.

24 31 2.8 384 4 AAI34569 Probe #32
25 31 2.8 384 4 ABA44472 Human bre
26 31 2.8 384 4 ABA24683 Probe #31
27 31 2.8 384 4 AAK28643 Human bon
28 31 2.8 384 4 AAK03190 Human bra
29 31 2.8 384 4 ABE28239 Human liv
30 31 2.8 384 5 AAI03121 Probe #31
31 31 2.8 384 6 ABE03152 Human gen
32 31 2.8 384 6 ABE03152 Human gen
33 31 2.8 1728 3 AAK32174 Oryzias l
34 31 2.8 1863 3 AAK32175 Oryzias l
35 31 2.8 2764 6 AAD31046 Medaka fi
36 31 2.8 6115 3 AAZ46519 Murine CA
37 30 2.7 69 3 AAC11384 Human sec
38 30 2.7 123 4 ABA70569 Human sec
39 30 2.7 123 4 AAI50733 Probe #19
40 30 2.7 123 4 AAK44756 Human bon
41 30 2.7 123 4 AAK18816 Human bra
42 30 2.7 123 6 ABE18995 Human liv
43 30 2.7 135 5 AAS80508 DNA enco
44 30 2.7 207 4 AAI24964 Probe #14
45 30 2.7 207 4 ABA70492 Human foe

PS	Claim 9; Page 97; 252pb; English.	
XX	This invention relates to a new isolated G-protein coupled receptor	
CC	(GPCR) polypeptide sequence and the cDNA encoding it. The GPCR	
CC	polypeptide, GPCR nucleic acid and an antibody specific to the protein	
CC	are useful for treating, preventing or alleviating a GPCR-associated	
CC	disorder or a pathological state in a subject, particularly a human. In	
CC	particular, the disorder is cardiomyopathy, atherosclerosis, diabetes, or	
CC	a disorder related to cell signal processing and metabolic pathway	
CC	modulation. The GPCR polypeptide and nucleic acid are also useful for	
CC	diagnosing the presence of or predisposition to a disease associated with	
CC	altered levels of GPCR, particularly cancer. The GPCR nucleic acid and	
CC	polypeptide are especially useful in the manufacture of a medicament for	
CC	therapeutic or prophylactic applications for disorders associated with	
CC	aberrant GPCR expression or activity, e.g. Von Hippel-Lindau syndrome,	
CC	Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia,	
CC	Parkinson's disease, Huntington's disease, cerebellar palsy, epilepsy,	
CC	Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia,	
CC	leukodystrophies, addiction, anxiety, depression, pain, obesity, Crohn's	
CC	disease, osteoporosis, inflammatory bowel disease, infertility,	
CC	hypertension, scleroderma, haemophilia, asthma, arthritis, human	
CC	immunodeficiency virus, autoimmune disease; HIV, viral, fungal, bacterial	
CC	or protozoal infections, or graft-versus-host disease. The DNA encoding	
CC	the protein is useful in gene therapy for treating the above conditions.	
CC	The polypeptides can be used as immunogens to produce antibodies and as	
CC	vaccines. The nucleic acids are further used as hybridisation probes, in	
CC	chromosome mapping, tissue typing, preventive medicine, and	
CC	pharmacogenomics. These are also useful in developing powerful assay	
CC	system for functional analysis of various human disorders, as well as in	
CC	diagnostic applications. The present sequence represents the cDNA	
CC	encoding a human G protein coupled receptor related protein (GPCR) of the	
XX	invention	
SQ	Sequence 1104 BP; 217 A; 328 C; 290 G; 269 T; 0 U; 0 Other;	
	Query Match 100.0%; Score 1104; DB 6; Length 1104;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 1104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 ATGATAAAGTGGGCGCTTACTTACACCCACCCACAAATAAATAAAGATTAAAGAT 60	
DB	1 ATGATAAAGTGGGCGCTTACTTACACCCACCCACAAATAAATAAAGATTAAAGAT 60	
QY	61 TTAAATGTGGAAAGCAAACTGTGAAATTTTTTGGCGCTCTCTGTCTACAGCCCTC 120	
DB	61 TTAAATGTGGAAAGCAAACTGTGAAATTTTTTGGCGCTCTCTGTCTACAGCCCTC 120	
QY	121 CTGGTCTTCAGTGGCGCTGGGAAACATCTTGGCGCTTGGCTTACCTGTCAAAGCAGG 180	
DB	121 CTGGTCTTCAGTGGCGCTGGGAAACATCTTGGCGCTTGGCTTACCTGTCAAAGCAGG 180	
QY	181 AAGATCAACTGCACAGGCATCTACCTGTGCACCTGGCTGTGTCTGACCTGCTTCAAC 240	
DB	181 AAGATCAACTGCACAGGCATCTACCTGTGCACCTGGCTGTGTCTGACCTGCTTCAAC 240	
QY	241 GTGGCGCTTACCGGGAAGGCTGTGTGTATGTGTGGCTCCAGCTGGCCTTGGGCAAG 300	
DB	241 GTGGCGCTTACCGGGAAGGCTGTGTGTATGTGTGGCTCCAGCTGGCCTTTCGGCAAG 300	
QY	301 GGGCTCTGCAGGCTCACGGCGTTTGTGTCTACACCGACACTACGGGGGGGTCTACCTC 360	
DB	301 GGGCTCTGCAGGCTCACGGCGTTTGTGTCTACACCGACACTACGGGGGGGTCTACCTC 360	
QY	361 ATGGCGCTGTGAGCGTGGACCAATPACCCAGCTGTGTGTGTGTGTGTGTGTGTGTGT 420	
DB	361 ATGGCGCTGTGAGCGTGGACCAATPACCCAGCTGTGTGTGTGTGTGTGTGTGTGTGT 420	
QY	421 CTCGCCAGCGTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480	
DB	421 CTCGCCAGCGTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480	
QY	481 CAGACGATGCCCTTCTCTTGTATGCCCATGACCAAGCCGCTGTGGGCAAGCTGGCGTGC 540	

DB	481 CAGACGATGCCCTTCTCTTGTATGCCCATGACCAAGCCGCTGTGGGCAAGCTGGCGTGC 540	
QY	541 ATGGAGTACAGCAGCATCGAGTCAAGTCTCGGGCTTGGCCCTCATGGTCTCTGTGGCTTTT 600	
DB	541 ATGGAGTACAGCAGCATCGAGTCAAGTCTCGGGCTTGGCCCTCATGGTCTCTGTGGCTTTT 600	
QY	601 GCATTGGCTTCTGTGGCCAGTGGGATCATCTGTCTCTGTATATGAAGTACACCTGG 660	
DB	601 GCATTGGCTTCTGTGGCCAGTGGGATCATCTGTCTCTGTATATGAAGTACACCTGG 660	
QY	661 AAGCTGTGCAGCACACAGCTCGGAGGACCCAGTACAGCAGGAAAGACGACCTGGCGA 720	
DB	661 AAGCTGTGCAGCACACAGCTCGGAGGACCCAGTACAGCAGGAAAGACGACCTGGCGA 720	
QY	721 GGCCTGCTGTATACCTGCTGCTGCTGTGTGGCTGTGTGTGTGTGTGTGTGTGTGTGT 780	
DB	721 GGCCTGCTGTATACCTGCTGCTGCTGTGTGGCTGTGTGTGTGTGTGTGTGTGTGTGT 780	
QY	781 CTCACATCAACAGCTTCTGCGAGAGGATGCTCCACCTGCCATCTGTGCCGAGCGG 840	
DB	781 CTCACATCAACAGCTTCTGCGAGAGGATGCTCCACCTGCCATCTGTGCCGAGCGG 840	
QY	841 AGGGCTTTTCTACTCTCCCTTCAAGCCACCGTGGCCCTCATGAACATGAATGTGGCAT 900	
DB	841 AGGGCTTTTCTACTCTCCCTTCAAGCCACCGTGGCCCTCATGAACATGAATGTGGCAT 900	
QY	901 ACCCAATCATTTACTTCTTTGATCCACCCATACAGGAAATGGCTCTCTGGCATTTTA 960	
DB	901 ACCCAATCATTTACTTCTTTGATCCACCCATACAGGAAATGGCTCTCTGGCATTTTA 960	
QY	961 AAGCTCAAAGGGTCTTCT 1020	
DB	961 AAGCTCAAAGGGTCTTCT 1020	
QY	1021 TCAGAAACACCAAGTATCACCCGCGCAGAGGTTCTATGTTCTTAGCGGACACCTGGTC 1080	
DB	1021 TCAGAAACACCAAGTATCACCCGCGCAGAGGTTCTATGTTCTTAGCGGACACCTGGTC 1080	
QY	1081 TAACGTGTCAATTGCTTTATGACT 1104	
DB	1081 TAACGTGTCAATTGCTTTATGACT 1104	
RESULT 2		
ADCS6840		
ID	ADCS6840 standard; DNA; 4319 BP.	
XX	AC ADC6840;	
XX	XX 01-JAN-2004 (first entry)	
DT	Human GPCR gene SEQ ID NO:1293.	
DE	de; gene; human; GPCR;	
KW	guanosine triphosphate-binding protein coupled receptor; gene therapy.	
XX	Homo sapiens.	
XX	EP1270724-A2.	
PN	02-JAN-2003.	
XX	18-JUN-2002; 2002EP-00013517.	
PF	18-JUN-2001; 2001JP-00246789.	
XX	(NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.	
PA	(ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.	
XX	Suwa M, Asai K, Akiyama Y, Aburatani H;	
PI	WEI; 2003-315783/31.	
XX	P-PSDB; ADC6841.	
DR		

XX New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
XX
PS Claim 1; SEQ ID NO 1293; 28pp; English.
XX
XX The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ABC8548-ADC87616 encode GPCR's of the
XX invention.
XX
SQ Sequence 4319 BP; 1058 A; 1049 C; 1170 G; 994 T; 0 U; 48 Other;

Query Match 41.8%; Score 461; DB 9; Length 4319;
Best Local Similarity 99.8%; Pred. No. 9.4e-217;
Matches 581; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 97 CGCTCTGTTCTACACAGCGCTCTCTGTTCTTCAAGTCCCTGGGAAACATCCTTGCCCTT 156
DB 2940 CGCTCTGTTCTACACAGCGCTCTCTGTTCTTCAAGTCCCTGGGAAACATCCTTGCCCTT 2999
QY 157 TGCCTTACCTGTAAGAGCAGGAGATCACTGTCAGCATCTACCTGGTGCACCTG 216
DB 3000 TGCCTTACCTGTAAGAGCAGGAGATCACTGTCAGCATCTACCTGGTGCACCTG 3059
QY 217 GCTGTGCTGACCTGCTGTTCAACGCTGGCTTACCGGAGGGTGGTGTATGTGCTG 276
DB 3060 GCTGTGCTGACCTGCTGTTCAACGCTGGCTTACCGGAGGGTGGTGTATGTGCTG 3119
QY 277 GGCTCCAGCTGGCTTTGCGCAAGGGCTGTGAGGCTGACGGGTTTGTCTACACC 336
DB 3120 GGCTCCAGCTGGCTTTGCGCAAGGGCTGTGAGGCTGACGGGTTTGTCTACACC 3179
QY 337 GACACCTACGGGGGGTCTACTCATGCGCTGTGAGCGTGACCATACCCAGCTGTG 396
DB 3180 GACACCTAC-GGGGGGTCTACTCATGCGCTGTGAGCGTGACCATACCCAGCTGTG 3238
QY 397 GTCTGTGCCCCACTGGGGCCCGCTCCGACAGGCTGGCCGCGCAGGCTGGTCTGCGTG 456
DB 3239 GTCTGTGCCCCACTGGGGCCCGCTCCGACAGGCTGGCCGCGCAGGCTGGTCTGCGTG 3298
QY 457 GCCATCTGACCTTGGTGTCTCTGACAGCATGCGCTGTCTGTGATGCCCATGACCAAG 516
DB 3299 GCCATCTGACCTTGGTGTCTCTGACAGCATGCGCTGTCTGTGATGCCCATGACCAAG 3358
QY 517 CCGCTGTGGGCAAGCTGGCTGTCATGAGTACAGCAGCATGGAGTCACTCTCGGGCTG 576
DB 3359 CCGCTGTGGGCAAGCTGGCTGTCATGAGTACAGCAGCATGGAGTCACTCTCGGGCTG 3418
QY 577 CCCCTCATGCTCTGGTGGCTTTGCCATTGGCTTGTGGGCCAGTGGGGATCATCTTG 636
DB 3419 CCCCTCATGCTCTGGTGGCTTTGCCATTGGCTTGTGGGCCAGTGGGGATCATCTTG 3478
QY 637 TCCTGTATATGAAGATCACTGGAAGCTGTGACGACAGCT 678
DB 3479 TCCTGTATATGAAGATCACTGGAAGCTGTGACGACAGCT 3520

RESULT 3
AAK76477
ID AAK76477 standard; DNA; 9409 BP.
XX
XX AAK76477;
AC AAK76477;
XX
XX
DT 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:31289.
XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226688P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0228287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 05-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.
XX 14-SEP-2000; 2000US-0232399P.
XX 14-SEP-2000; 2000US-0232400P.
XX 14-SEP-2000; 2000US-0232401P.
XX 14-SEP-2000; 2000US-0233063P.
XX 14-SEP-2000; 2000US-0233064P.
XX 14-SEP-2000; 2000US-0233065P.
XX 21-SEP-2000; 2000US-0234223P.
XX 21-SEP-2000; 2000US-0234274P.

QY 637 TCCTCTATATGAAGATCACCTGGAGCTGTGCACACAGCT 678
|||||
Db 912 TCCTCTATATGAAGATCACCTGGAGCTGTGCACACAGCT 953
|||||

RESULT 4

AAS98131/c
ID AAS98131 standard; cDNA; 1001 BP.

XX AC AAS98131;
XX AC

DT 12-MAR-2002 (first entry)

XX DE Human DNA for potential G protein-coupled receptor #88.

XX Human; G protein-coupled receptor; GPCR; ss; GAL4; galanin receptor;
XX Alzheimer's disease; amyotrophic lateral sclerosis; asthma;
XX attherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;
XX chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;
XX depression; epilepsy; macular degeneration; lymphoma; melanoma;
XX multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;
XX psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;
XX tuberculosis; cognition disorder; memory disorder; anorexia;
XX hormonal release disorder; cardiovascular activity disorder;
XX pain perception disorder; obesity; diabetes; diabetes; obesity; diabetes;
XX hyperlipidaemia; stroke; gene therapy.

XX OS Homo sapiens.

XX WO200185791-A1.

XX 15-NOV-2001.

XX 11-MAY-2001; 2001WO-US015332.

XX 11-MAY-2000; 2000US-0203217P.

XX 18-MAY-2000; 2000US-0205945P.

XX (LIFE-) LIPESAN BIOSCIENCES INC.

PI Brown JP, Miller M, Burmer G, Fabre-Suver C, Pritchard D;

XX WPI; 2002-066595/09.

XX Novel G protein-coupled receptor polypeptides including galanin receptor
PT polypeptides useful for identifying modulators that are useful for
PT treating Alzheimer's disease, psoriasis, melanoma, multiple sclerosis,
PT stroke.

XX Claim 2; Page 95; 14app; English.

XX The invention relates to an isolated polypeptide encoded by a nucleic
CC acid molecule that is at least 80% identical to the G protein-coupled
CC (GPCR) polynucleotides included in the specification. Also included are
CC probes based on the GPCR sequences (including antisense probes), a host
CC cell comprising an expression vector comprising the GPCR sequence,
CC antibodies raised against the polypeptides, and methods of identifying
CC modulators of the polypeptides. The polypeptides are useful for
CC identifying modulator compounds which function as modulators, activators,
CC repressors, agonists or antagonists of the novel GPCR polypeptides
CC including the GAL4 polypeptide. The antibodies and nucleic acid probes as
CC described above can be used to detect the presence of the polypeptides
CC and nucleic acids and are used to diagnose a variety of diseases or
CC disorders in which GPCRs are involved e.g., Alzheimer's disease,
CC amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell
CC carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, chronic
CC obstructive pulmonary disease, Crohn's disease, depression, epilepsy,
CC macular degeneration, lymphoma, melanoma, multiple sclerosis, rheumatoid
CC osteoarthritis, osteoporosis, Parkinson's disease, psoriasis, rheumatoid
CC arthritis, schizophrenia, ulcerative colitis, tuberculosis and many other
CC diseases listed in the specification. The probes and antibodies are also
CC useful for diagnosing cognition and memory disorders, anorexia, hormonal
CC release disorders, cardiovascular activity disorders, pain perception

CC disorders, obesity, diabetes, Alzheimer's disease. Preferably, compounds
CC that decrease or increase the expression of galanin receptor (GAL4) can
CC be used to treat obesity, diabetes, hyperlipidaemia and stroke. The GPCR
CC nucleic acid is useful for treating the above mentioned disorders by gene
CC therapy techniques. The present sequence is a novel GPCR polynucleotide
CC of the invention

XX SQ Sequence 1001 BP; 190 A; 315 C; 311 G; 185 T; 0 U; 0 Other;

Query Match 41.5%; Score 458; DB 6; Length 1001;
Best Local Similarity 99.8%; Pred. No. 2.8e-215;
Matches 578; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 100 TCCTCTATACACAGCCCTCTCTGCTTCAGTCGCCCTGGGAAACATCTCTGCCCTTGC 159
|||||
Db 1001 TCCTCTATACACAGCCCTCTCTGCTTCAGTCGCCCTGGGAAACATCTCTGCCCTTGC 942
|||||
QY 160 CTTTACCTGTCAAAAGACAGGAGATCAACTGCACAGGATCTTACCTGGTGACCTGGCT 219
|||||
Db 941 CTTTACCTGTCAAAAGACAGGAGATCAACTGCACAGGATCTTACCTGGTGACCTGGCT 882
|||||
QY 220 GTCTCTGACCTGCTGTTCACCTGGCTTACCGGGAAGGCTGTGTATGTCTGGGC 279
|||||
Db 881 GTCTCTGACCTGCTGTTCACCTGGCTTACCGGGAAGGCTGTGTATGTCTGGGC 822
|||||
QY 280 TCCAGCTGGCCTTTCCGCAAGGGCTCTCGAGCTCAGCGCTTCTGTCTACACCGAC 339
|||||
Db 821 TCCAGCTGGCCTTTCCGCAAGGGCTCTCGAGCTCAGCGCTTCTGTCTACACCGAC 762
|||||
QY 340 ACCTACGGGGGCTCTACCTCATGGCTGTGTGAGCGTGGACCAATPACCCAGCTGGTC 399
|||||
Db 761 ACCTAC-GGGGGGCTCTACCTCATGGCTGTGTGAGCGTGGACCAATPACCCAGCTGGTC 703
|||||
QY 400 TGTGCCCACTGGGGCCCGCGCTCCGACGGCTGGCGCGCCAGGCTGTCTGCGTGGCC 459
|||||
Db 702 TGTGCCCACTGGGGCCCGCGCTCCGACGGCTGGCGCGCCAGGCTGTCTGCGTGGCC 643
|||||
QY 460 ATCTGACCTTGTGTCTGTGTGCAGCATGCCCTTCTCTTGATGCCCATGACCAAGCCG 519
|||||
Db 642 ATCTGACCTTGTGTCTGTGTGCAGCATGCCCTTCTCTTGATGCCCATGACCAAGCCG 583
|||||
QY 520 CTGCTGGGCAAGCTGGCCCTGCATGGAGTACAGCAGCATGAGTCAGTCTCGGGTCCCC 579
|||||
Db 582 CTGCTGGGCAAGCTGGCCCTGCATGGAGTACAGCAGCATGAGTCAGTCTCGGGTCCCC 523
|||||
QY 580 CTGATGCTCTGTGGCTTTGGCATTTGGCTTCTGTGGCCAGTGGGGATCATCTGTGCC 639
|||||
Db 522 CTGATGCTCTGTGGCTTTGGCATTTGGCTTCTGTGGCCAGTGGGGATCATCTGTGCC 463
|||||
QY 640 TGCTATATGAAGATCACCTGGAAAGCTGTGAGCAGCAGCT 678
|||||
Db 462 TGCTATATGAAGATCACCTGGAAAGCTGTGAGCAGCAGCT 424
|||||

RESULT 5

ABK94922

ID ABK94922 standard; cDNA; 1526 BP.

XX AC ABK94922;

XX AC

DT 30-AUG-2002 (first entry)

XX DE Human novel polynucleotide #33.

XX Human; gene; ss; inflammatory condition; shock; sepsis; immune response;
XX cancer; wound healing; central nervous system disease; haematopoiesis;
XX peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
XX myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
XX cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
XX bone degenerative disorder; periodontal disease; reperfusion injury;
XX lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
XX allergic condition; thrombolysis; thrombosis; coagulation disorder;
XX fungal infection.

XX OS Homo sapiens.
 XX PN WO200244340-A2.
 XX PD 06-JUN-2002.
 XX PF 30-NOV-2001; 2001WO-US047004.
 XX PR 30-NOV-2000; 2000US-00728952.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
 PI Yamazaki V, Ujwal ML, Drmanac RT;
 XX WPI: 2002-508509/54.
 XX DR P-PSDB; ABG56698.
 XX PT Novel nucleic acids and polypeptides for diagnosis, treatment of
 PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
 PT disorders, cancer and promoting wound healing.
 XX PS Claim 1; Page 403-404; 672pp; English.
 XX CC The invention relates to human novel polynucleotides and associated
 CC polypeptides. The polynucleotides and polypeptides are useful for
 CC treating inflammatory conditions such as arthritis, nephritis, Crohn's
 CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
 CC and cancer and for promoting wound healing. The sequences are used to
 CC induce the proliferation of neural cells and regeneration of nerve and
 CC brain tissue, and are useful for the treatment of central and peripheral
 CC nervous system diseases and neuropathies, such as Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease and amyotrophic lateral
 CC sclerosis. The sequences are involved in chemotactic or chemokinetic
 CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
 CC cell disorders and platelet disorders such as thrombocytopenia,
 CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
 CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
 CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
 CC disease. The sequences of the invention are also useful for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues, immune deficiencies and disorders
 CC including severe combined immunodeficiency (SCID), bacterial or fungal
 CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
 CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis
 CC and coagulation disorders. Sequences ABK94890-ABK94982 represent human
 CC novel polynucleotides of the invention.
 XX SQ Sequence 1526 BP; 282 A; 458 C; 458 G; 328 T; 0 U; 0 Other;

Query Match 37.0%; Score 409; DB 6; Length 1526;
 Best Local Similarity 99.7%; Pred. No. 4.1e-191;
 Matches 579; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 98 GCTCTGTGTTACACAGCCCTCTGTTCTTCTAGTGCCTGGGAACATCTTGCCTTT 157
 DB 422 GCTCTGTGTTACACAGCCCTCTGTTCTTCTAGTGCCTGGGAACATCTTGCCTTT 481

QY 158 GCCTTACCTGTCAAGAGCAGGAGATCACTGCACAGGATCTACTCTGTTGCACCTGG 217
 DB 482 GCCTTACCTGTCAAGAGCAGGAGATCACTGCACAGGATCTACTCTGTTGCACCTGG 541

QY 218 CTGTGTTCTGACCTGTTCTACCTGCTTACCTGCTTACCTGCTTACCTGCTTACCTG 277
 DB 542 CTGTGTTCTGACCTGTTCTACCTGCTTACCTGCTTACCTGCTTACCTGCTTACCTG 601

QY 278 GCTCCAGCTGACCTTGGCAAGGGCTCTGCAGGCTGACGGCTTGTCTCTACACCG 337
 DB 602 GCTCCAGCTGACCTTGGCAAGGGCTCTGCAGGCTGACGGCTTGTCTCTACACCG 661

QY 338 ACACCTACGGGGGCTCTACCTGCTTACCTGCTTACCTGCTTACCTGCTTACCTGCTG 397

662 ACACCTAC-GGGGGGTCTACCTCATGCGCTGTGTGAGCGTGGACCACTACCCAGCTGTGG 720
 QY TCTGTGCCCCACTGGGGCCCGCCCTCCGACAGGCTGGCCGCGCCAGGCTGGTCTGGGTGG 457
 DB 721 TCTGTGCCCCACTGGGGCCCGCCCTCCGACAGGCTGGCCGCGCCAGGCTGGTCTGGGTGG 780
 QY 458 CCATCTGGACCTTGGTGTCTGCTGCACAGATGCCCTTGTCTTGTGATGCCCATGACCAAGC 517
 DB 781 CCATCTGGACCTTGGTGTCTGCTGCACAGATGCCCTTGTCTTGTGATGCCCATGACCAAGC 840
 QY 518 CGCTGGTGGCAAGCTGGCCCTGCATGGAGTACAGCAGCAGTGGAGTCTCCTCGGGGTGC 577
 DB 841 CGCTGGTGGCAAGCTGGCCCTGCATGGAGTACAGCAGCAGTGGAGTCTCCTCGGGGTGC 900
 QY 578 CCCTCATGCTCTGCTGGCCCTTTCCTTCCATGGCTTCTGTGGCCAGTGGGATCATCTGT 637
 DB 901 CCCTCATGCTCTGCTGGCCCTTTCCTTCCATGGCTTCTGTGGCCAGTGGGATCATCTGT 960
 QY 638 CTTGCTATATGAAGATCACCTGGAAGCTGTGACGACAGCT 678
 DB 961 CTTGCTATATGAAGATCACCTGGAAGCTGTGACGACAGCT 1001

RESULT 6
 AAS98111
 ID AAS98111 standard; DNA; 330 BP.
 XX AC AAS98111;
 XX DT 12-MAR-2002 (first entry)
 XX DE Human DNA for potential G protein-coupled receptor #68.
 XX KW Human; G protein-coupled receptor; GPCR; ds; GAL4; galanin receptor;
 KW Alzheimer's disease; amyotrophic lateral sclerosis; asthma;
 KW atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;
 KW chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;
 KW depression; epilepsy; macular degeneration; lymphoma; melanoma;
 KW multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;
 KW psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;
 KW tuberculosis; cognition disorder; memory disorder; anorexia;
 KW hormonal release disorder; cardiovascular activity disorder;
 KW pain perception disorder; obesity; diabetes; obesity; diabetes;
 KW hyperlipidaemia; stroke; gene therapy.
 XX OS Homo sapiens.
 XX PN WO200185791-A1.
 XX PD 15-NOV-2001.
 XX PF 11-MAY-2001; 2001WO-US015332.
 XX PR 11-MAY-2000; 2000US-0203217P.
 XX PR 18-MAY-2000; 2000US-0205945P.
 XX PA (LIFE-) LIFESPAN BIOSCIENCES INC.
 XX PI Brown JP, Miller M, Burmer G, Fabre-Suver C, Pritchard D;
 XX WPI: 2002-066595/09.
 XX PT Novel G protein-coupled receptor polypeptides including galanin receptor
 PT polypeptides useful for identifying modulators that are useful for
 PT treating Alzheimer's disease, psoriasis, melanoma, multiple sclerosis,
 PT stroke.
 XX PS Claim 2; Page 83; 144pp; English.
 XX CC The invention relates to an isolated polypeptide encoded by a nucleic
 CC acid molecule that is at least 80% identical to the G protein-coupled
 CC (GPCR) polynucleotides included in the specification. Also included are
 CC probes based on the GPCR sequences (including antisense probes), a host

cell comprising an expression vector comprising the GPCR sequence, antibodies raised against the polypeptides, and methods of identifying modulators of the polypeptides. The polypeptides are useful for identifying modulator compounds which function as modulators, activators, repressors, agonists or antagonists of the novel GPCR polypeptides including the GAL4 polypeptide. The antibodies and nucleic acid probes as described above can be used to detect the presence of the polypeptides and nucleic acids and are used to diagnose a variety of diseases or disorders in which GPCRs are involved e.g., Alzheimer's disease, amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, chronic obstructive pulmonary disease, Crohn's disease, depression, epilepsy, macular degeneration, lymphoma, melanoma, multiple sclerosis, osteoarthritis, osteoporosis, Parkinson's disease, psoriasis, rheumatoid arthritis, schizophrenia, ulcerative colitis, tuberculosis and many other diseases listed in the specification. The probes and antibodies are also useful for diagnosing cognition and memory disorders, anorexia, hormonal release disorders, cardiovascular activity disorders, pain perception disorders, obesity, diabetes, Alzheimer's disease. Preferably, compounds that decrease or increase the expression of galanin receptor (GAL4) can be used to treat obesity, diabetes, hyperlipidaemia and stroke. The GPCR nucleic acid is useful for treating the above mentioned disorders by gene therapy techniques. The present sequence is a novel GPCR polynucleotide of the invention

Sequence 330 BP; 46 A; 101 C; 108 G; 75 T; 0 U; 0 Other;

Query Match 28.8%; Score 318; DB 6; Length 330;
Best Local Similarity 100.0%; Pred. No. 3.1e-146;
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 GGGGTCTACCTCAGTGGTGTGAGCGTGACCATACCCAGCTGTGGTGTGGCCAC 408
Db 1 GGGGTCTACCTCAGTGGTGTGAGCGTGACCATACCCAGCTGTGGTGTGGCCAC 60
QY 409 TGGGGCCCGCGCTCCGACCGCTGGCGCGCCAGGTGTGTGCGTGGCACTGGACC 468
Db 61 TGGGGCCCGCGCTCCGACCGCTGGCGCGCCAGGTGTGTGCGTGGCACTGGACC 120
QY 469 TTGTGTGCTGTGACGATCCCTCTCTTGTATGCCATGACCAAGCCGCTGTGGGC 528
Db 121 TTGTGTGCTGTGACGATCCCTCTCTTGTATGCCATGACCAAGCCGCTGTGGGC 180
QY 529 AAGTGGCGCTGCATGGAGTACAGCAGCATGGAGTCACTCTCGGGCTGCCCTCATGGTC 588
Db 181 AAGTGGCGCTGCATGGAGTACAGCAGCATGGAGTCACTCTCGGGCTGCCCTCATGGTC 240
QY 589 CTGTGGCGCTTGGCATGGCTTCTGTGGGCGAGTGGGATCATCTCTGCTATATG 648
Db 241 CTGTGGCGCTTGGCATGGCTTCTGTGGGCGAGTGGGATCATCTCTGCTATATG 300
QY 649 AAGATCACTGGAGCTG 666
Db 301 AAGATCACTGGAGCTG 318

RESULT 7
AAS78869
ID AAS78869 standard; cDNA; 1771 BP.
XX AAS78869;
AC AAS78869;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #14673.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

XX WO200175067-A2.

PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG14682.

XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 1; SEQ ID NO 14673; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pcr_sequences

XX Sequence 1771 BP; 313 A; 531 C; 519 G; 407 T; 0 U; 1 Other;

Query Match 25.3%; Score 279; DB 5; Length 1771;
Best Local Similarity 99.4%; Pred. No. 5.4e-127;
Matches 439; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 178 AGAAGATCACTGCACAGGCATCTACCTGTGACCTGTGCTGTGACCTGTGTC 237
Db 742 AGAAGATCACTGCACAGGCATCTACCTGTGACCTGTGCTGTGACCTGTGTC 801
QY 238 ACCGTGGCGCTTACCGGGAAGGCTGTGTATGTCTGGGCTCCAGCTGGCTTTTCGC 297
Db 802 ACCGTGGCGCTTACCGGGAAGGCTGTGTATGTCTGGGCTCCAGCTGGCTTTTCGC 861
QY 298 AAGGGGCTCTGAGGCTGACGGCGTTTGTCTTACCGACACCTACGGGGGGTCTAC 357
Db 862 AAGGGGCTCTGAGGCTGACGGCGTTTGTCTTACCGACACCTACGGGGGGTCTAC 921
QY 358 CTCATGGCGCTGTGAGCGTGACCATACCCAGC-TGTGTCTGTGCCACTTGGGGGCC 416
Db 922 CTCATGGCGCTGTGAGCGTGACCATACCCAGC-TGTGTCTGTGCCACTTGGGGGCC 981
QY 417 GCGCTTCCGACGGCTGGCGGCCAGGCTGTCTCGTGGCCATCTGGACCTTGTGT 476
Db 982 GTGCTTCCGACGGCTGGCGGCCAGGCTGTCTCGTGGCCATCTGGACCTTGTGT 1041
QY 477 GCTGCAGAGATGCCCTTGTCTTGTATGCCATGACCAAGCCGCTGTGGGCAAGTGGC 536
Db 1042 GCTGCAGAGATGCCCTTGTCTTGTATGCCATGACCAAGCCGCTGTGGGCAAGTGGC 1101

QY 537 CTGCATGGAGTACACAGCATGGAGTCAGTCTCTGGGCTGCCCTCATGGTCTCTGGTGGC 596
 Db 1102 CTGCATGGAGTACACAGCATGGAGTCAGTCTCTGGGCTGCCCTCATGGTCTCTGGTGGC 1161
 QY 597 CTTTGCATTGCTTCTGTGGCCAGTGGGATCATCTCTGCTCTATATGAAGATCAC 656
 Db 1162 CTTTGCATTGCTTCTGTGGCCAGTGGGATCATCTCTGCTCTATATGAAGATCAC 1221
 QY 657 CTGGAAGCTGTGCACACAGCT 678
 Db 1222 CTGGAAGCTGTGCACACAGCT 1243

RESULT 8

AAS42840
 ID AAS42840 standard; cDNA; 536 BP.

XX AAS42840;

XX 18-DEC-2001 (first entry)

XX Human G Protein-Coupled Receptor (GPCR) cDNA #35.

XX Human; G-protein coupled receptor; GPCR; mental disorder; schizophrenia;
 KW attention deficit disorder; anxiety; depression; bipolar disorder; ss;
 KW neurological disorder; Huntington's disease; dementia; obesity; anorexia;
 KW metabolic disorder; Parkinson's disease; Tourette's syndrome; thrombosis;
 KW type 2 diabetes; cardiovascular disorder; myocardial infarction; cancer;
 KW cardiomyopathy; atherosclerosis; human immunodeficiency virus; HIV;
 KW viral infection; immunostimulant; neuroleptic; nootropic; tranquiliser;
 KW antidepressant; anorectic; PCR primer; gene therapy.

XX Homo sapiens.

XX WO200162797-A2.

XX 30-AUG-2001.

XX 23-FEB-2001; 2001WO-US005676.

XX 23-FEB-2000; 2000US-0184247P.

XX 23-FEB-2000; 2000US-0184303P.

XX 23-FEB-2000; 2000US-0184304P.

XX 23-FEB-2000; 2000US-0184305P.

XX 23-FEB-2000; 2000US-0184397P.

XX 02-MAR-2000; 2000US-0186457P.

XX 03-MAR-2000; 2000US-0186810P.

XX 09-MAR-2000; 2000US-0188064P.

XX 13-MAR-2000; 2000US-0188880P.

XX 03-APR-2000; 2000US-0194344P.

XX 23-JUN-2000; 2000US-0213861P.

XX 11-JUL-2000; 2000US-0217369P.

XX 14-JUL-2000; 2000US-0217370P.

XX 20-JUL-2000; 2000US-0218337P.

XX 20-JUL-2000; 2000US-0218492P.

XX (PHAA) PHARMACIA & UPJOHN CO.

XX Vogeli G, Wood LS, Parodi LA, Lind P;

XX WPI; 2001-570628/64.

XX P-PSDB; AAU25588.

XX New isolated nucleic acid encoding a new G-protein coupled receptor

XX polypeptide for detecting receptor modulators that can treat mental

XX disorders, such as schizophrenia, anxiety, depression, or obesity.

XX Claim 4; Page 82-83; 279pp; English.

XX Sequences AAS42806-AAS42926 represent cDNA molecules and PCR primers for

XX cDNA molecules encoding human G-protein coupled receptor (GPCR)

XX polypeptides. The protein and DNA sequences of the invention can be used

XX to identify compounds which bind to GPCR polypeptides and in screening

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC for compounds that modulate GPCR activity. By screening a human subject
 CC for the presence of mutations in GPCR DNA, a GPCR-related disorder or a
 CC genetic predisposition can be diagnosed. The sequences can also be used
 CC for treatment and prevention of mental disorders such as schizophrenia,
 CC attention deficit disorder, anxiety, depression, dementia and bipolar
 CC disorder, neurological disorders such as Huntington's disease,
 CC Parkinson's disease and Tourette's syndrome, metabolic disorders such as
 CC obesity, anorexia and type 2 diabetes, cardiovascular disorders such as
 CC thrombosis, myocardial infarction, cardiomyopathy and atherosclerosis,
 CC viral infections caused by HIV and cancers

XX Sequence 536 BP; 93 A; 166 C; 182 G; 95 T; 0 U; 0 Other;

Query Match 15.9%; Score 176; DB 4; Length 536;

Best Local Similarity 99.6%; Pred. No. 3.4e-76;

Matches 226; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 347 GGGGGGCTACCTCATGGCTGTGTGAGCTGGACCATACCAGCTGTGGTCTGTGGCC 406

Db 69 GGGGGGCTACCTCATGGCTGTGTGAGCTGGACCATACCAGCTGTGGTCTGTGGCC 128

QY 407 ACTGGGGGCGCGCCCTCCGCAACGGCTGGCGGCCAGGCTGTCTCGTGGCCATCTGGA 466

Db 129 ACTGGGGGCGCGCTCCGCAACGGCTGGCGGCCAGGCTGTCTCGTGGCCATCTGGA 188

QY 467 CTTTGTGTCTGTGCAGACATGCCCTTCTCTTGTATGCCCATGACCAAGCCGCTGTGG 526

Db 189 CTTTGTGTCTGTGCAGACATGCCCTTCTCTTGTATGCCCATGACCAAGCCGCTGTGG 248

QY 527 GCAAGCTGGGCTGCATGGAGTACAGCAGCATGAGTCAGTCCCTCCGGG 573

Db 249 GCAAGCTGGGCTGCATGGAGTACAGCAGCATGAGTCAGTCCCTCCGGG 295

RESULT 9

AAS30791/C

ID AAS30791 standard; cDNA; 556 BP.

XX AAS30791;

XX 04-DEC-2001 (first entry)

XX Human cDNA encoding G protein-coupled receptor nGPCR-2343.

XX Human; G protein-coupled receptor; nGPCR-X; ss; antiviral; analgesic;

XX cytosolic; cardiac; antidiabetic; anorectic; hypotensive; hypertensive;

XX antiparkinsonian; nootropic; neuroprotective; antidepressant;

XX cancer; metabolic disease; cardiovascular disease; type 2 diabetes;

XX obesity; anorexia; hypotension; hypertension; myocardial infarction;

XX atherosclerosis; Parkinson's disease; psychosis; neurological disorder;

XX schizophrenia; migraine; major depression; anxiety; mental disorder;

XX manic depression; dyskinesia; Huntington's disease; Tourette's Syndrome.

XX Homo sapiens.

XX WO200166750-A2.

XX 13-SEP-2001.

XX 08-MAR-2001; 2001WO-US007322.

XX 08-MAR-2000; 2000US-0187581P.

XX 08-MAR-2000; 2000US-0187582P.

XX 08-MAR-2000; 2000US-0187714P.

XX 08-MAR-2000; 2000US-0187715P.

XX 08-MAR-2000; 2000US-0187825P.

XX 08-MAR-2000; 2000US-0187828P.

XX 08-MAR-2000; 2000US-0187829P.

XX 08-MAR-2000; 2000US-0187830P.

XX 08-MAR-2000; 2000US-0187833P.

XX 08-MAR-2000; 2000US-0187874P.

XX 08-MAR-2000; 2000US-0187928P.

PR	08-MAR-2000; 2000US-0187929P.	
PR	08-MAR-2000; 2000US-0187930P.	
PR	08-MAR-2000; 2000US-0188049P.	
PR	08-MAR-2000; 2000US-0189294P.	
XX	(PHAA) PHARMACIA & UPJOHN CO.	
PA		
XX		
PI	Vogeli G, Wood LS;	
XX		
DR	WPI; 2001-536778/59.	
DR	P-PSDB; AAU19222.	
XX		
PT	Isolated nucleic acid molecules encoding G protein-coupled receptors	
PT	termed nGPCR-x, useful in the treatment and diagnosis of viral	
PT	infections, cancers and mental disorders (e.g. Parkinson's disease and	
PT	schizophrenia).	
XX		
PS	Claim 4; Page 204-205; 336pp; English.	
XX		
CC	The invention relates to novel isolated nucleic acid molecules encoding G	
CC	protein-coupled receptors termed nGPCR-x. nGPCR-x polynucleotides,	
CC	polypeptides, and modulators may be used in the treatment of diseases and	
CC	conditions such as infections, such as viral infections caused by HIV-1	
CC	(human immunodeficiency virus) or HIV-2, pain, cancers, metabolic and	
CC	cardiovascular diseases and disorders (e.g., type 2 diabetes, obesity,	
CC	anorexia, hypotension, hypertension, myocardial infarction,	
CC	atherosclerosis), Parkinson's disease, and psychotic and neurological	
CC	disorders, including schizophrenia, migraine, major depression, anxiety,	
CC	mental disorder, manic depression, and dyskinesias, such as Huntington's	
CC	disease or Tourette's Syndrome and many other diseases and syndromes	
CC	listed in the specification. nGPCR-x polynucleotides and polypeptides, as	
CC	well as nGPCR-x modulators, may also be used in diagnostic assays for	
CC	such diseases or conditions. The present sequence encodes a G protein-	
CC	coupled receptor of the invention	
XX		
SQ	Sequence 556 BP; 138 A; 139 C; 151 G; 128 T; 0 U; 0 Other;	
	Query Match 14.4%; Score 159; DB 4; Length 556;	
	Best Local Similarity 99.5%; Pred. No. 8.2e-68;	
	Matches 209; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	697 AGCAGGAAGACGCCACTGGCGAGGTGGCTCTTACGCTGCTGATGCTGGTGGCCGTG 756	
Db		
	556 AGCAGGAAGACGCCACTGGCGAGGTGGCTCTTACGCTGCTGATGCTGGTGGCCGTG 497	
QY	757 GTGGTCTGCTTTCAGCCCCCTACACCTCAAATCAAGCAGTTTCATGCGAGAGGGATGCTC 816	
Db		
	496 GTGGTCTGCTTTCAGCCCCCTACACCTCAAATCAAGCAGTTTCATGCGAGAGGGATGCTC 437	
QY	817 CACTTGCACATCTGTGCGAGCGAGGGCTTTCTTACTGCCCTTCAGGCCACCGTGGCC 876	
Db		
	436 CACCTGTCATCTGTGCGAGCGAGGGCTTTCTTACTGCCCTTCAGGCCACCGTGGCC 377	
QY	877 CTCATGAACATGAACCTGTGGCATTACCCCA 906	
Db		
	376 CTCATGAACATGAACCTGTGGCATTACCCCA 347	
RESULT 10		
AAS98061/c		
ID	AAS98061 standard; DNA; 594 BP.	
XX		
XX	AAS98061;	
AC		
XX		
DT	12-MAR-2002 (first entry)	
XX		
DE	Human DNA for potential G protein-coupled receptor #18.	
XX		
KW	Human; G protein-coupled receptor; GPCR; ds; GAT4; galanin receptor;	
KW	Alzheimer's disease; amyotrophic lateral sclerosis; aschma;	
KW	atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;	
KW	chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;	
KW	depression; epilepsy; macular degeneration; lymphoma; melanoma;	

KW multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;
 KW psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;
 KW tuberculosis; cognition disorder; memory disorder; anorexia;
 KW hormonal release disorder; cardiovascular activity disorder;
 KW pain perception disorder; obesity; diabetes; obesity; diabetes;
 KW hyperlipidaemia; stroke; gene therapy.
 XX
 XX Homo sapiens.
 XX WO200185791-A1.
 XX
 XX 15-NOV-2001.
 PD
 PD
 XX
 XX 11-MAY-2001; 2001WO-US015332.
 PF
 XX
 XX 11-MAY-2000; 2000US-0203217P.
 PR
 PR 18-MAY-2000; 2000US-0205945P.
 XX
 XX (LIFE-) LIFESPAN BIOSCIENCES INC.
 PA
 PA Brown JP, Miller M, Burmer G, Fabre-Suver C, Pritchard D;
 PI WPI; 2002-066595/09.
 DR
 DR
 XX
 XX Novel G protein-coupled receptor polypeptides including galanin receptor
 PT polypeptides useful for identifying modulators that are useful for
 PT treating Alzheimer's disease, psoriasis, melanoma, multiple sclerosis,
 PT stroke.
 XX
 XX Disclosure; Page 123; 144pp; English.
 XX
 XX The invention relates to an isolated polypeptide encoded by a nucleic
 CC acid molecule that is at least 80% identical to the G protein-coupled
 CC (GPCR) polynucleotides included in the specification. Also included are
 CC probes based on the GPCR sequences (including antisense probes), a host
 CC cell comprising an expression vector comprising the GPCR sequence, a host
 CC antibodies raised against the polypeptides, and methods of identifying
 CC modulators of the polypeptides. The polypeptides are useful for
 CC identifying modulator compounds which function as modulators, activators
 CC repressors, agonists or antagonists of the novel GPCR polypeptides
 CC including the GAL4 polypeptide. The antibodies and nucleic acid probes as
 CC described above can be used to detect the presence of the polypeptides
 CC and nucleic acids and are used to diagnose a variety of diseases or
 CC disorders in which GPCRs are involved e.g., Alzheimer's disease,
 CC amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell
 CC carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, chronic
 CC obstructive pulmonary disease, Crohn's disease, depression, epilepsy,
 CC macular degeneration, lymphoma, melanoma, multiple sclerosis, osteo-
 CC arthrititis, osteoporosis, Parkinson's disease, psoriasis, rheumatoid
 CC arthritis, schizophrenia, ulcerative colitis, tuberculosis and many other
 CC diseases listed in the specification. The probes and antibodies are also
 CC useful for diagnosing cognition and memory disorders, anorexia, hormonal
 CC release disorders, cardiovascular activity disorders, pain perception
 CC disorders, obesity, diabetes, Alzheimer's disease. Preferably, compounds
 CC that decrease or increase the expression of galanin receptor (GAL4) can
 CC be used to treat obesity, diabetes, hyperlipidaemia and stroke. The GPCR
 CC nucleic acid is useful for treating the above mentioned disorders by gene
 CC therapy techniques. The present sequence is a novel GPCR polynucleotide
 CC of the invention
 XX
 XX Sequence 594 BP; 164 A; 149 C; 171 G; 110 T; 0 U; 0 Other;

```

Query Match          9.3%; Score 103; DB 6; Length 594;
Best Local Similarity 100.0%; Pred. No. 3.4e-40;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      290 CTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTTGCTCTTACACCGACACCTTACGGGG 349
      |||||
Db       530 CTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTTGCTCTTACACCGACACCTTACGGGG 471

QY      350 GGGTCTACCTCATGGGCTGTGAGCGTGGACCATTTACCCAGC 392

Db       470 GGGTCTACCTCATGGGCTGTGAGCGTGGACCATTTACCCAGC 428

```


OS	Homo sapiens.
XX	
FN	WO200157278-A2.
XX	
PD	09-AUG-2001.
XX	
PF	30-JAN-2001; 2001WO-US0006070.
XX	
PR	04-FEB-2000; 2000US-0180312P.
PR	26-MAY-2000; 2000US-0207456P.
PR	30-JUN-2000; 2000US-00608408.
PR	03-AUG-2000; 2000US-00632366.
PR	21-SEP-2000; 2000US-0234687P.
PR	27-SEP-2000; 2000US-0236359P.
PR	04-OCT-2000; 2000GB-00024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
DR	WFI; 2001-488901/53.
XX	
PT	Human genome-derived single exon nucleic acid probes useful for analyzing
PT	gene expression in human cervical epithelial cells.
XX	
PS	Claim 25; SEQ ID NO 12384; 487bp; English.
XX	
CC	The present invention relates to human single exon nucleic acid probes
CC	(SENPs). The present sequence is one such probe. The SENPs are derived
CC	from human HeLa cells. The SENPs can be used to produce a single exon
CC	microarray, which can be used for measuring human gene expression in a
CC	sample derived from human cervical epithelial cells. By measuring gene
CC	expression, the probes are therefore useful in grading and/or staging of
CC	diseases of the cervix, notably cervical cancer. Note: The sequence data
CC	for this patent did not form part of the printed specification, but was
CC	obtained in electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 232 BP; 12 A; 94 C; 5 G; 119 T; 0 U; 2 Other;
	Query Match 2.8%; Score 31; DB 4; Length 232;
	Best Local Similarity 100.0%; Pred. No. 0.00011;
	Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	975 TTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCACC 1005
Db	91 TTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCACC 121
RESULT 13	
ABA67527	ID ABA67527 standard; DNA; 232 BP.
XX	AC ABA67527;
XX	
DT	01-FEB-2002 (first entry)
XX	
DE	Human foetal liver single exon nucleic acid probe #15832.
XX	
KW	Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
OS	Homo sapiens.
XX	
FN	WO200157277-A2.
XX	
PD	09-AUG-2001.
XX	
PF	30-JAN-2001; 2001WO-US000669.
XX	
PR	04-FEB-2000; 2000US-0180312P.
PR	26-MAY-2000; 2000US-0207456P.
PR	30-JUN-2000; 2000US-00608408.
PR	03-AUG-2000; 2000US-00632366.

PS Claim 25; SEQ ID NO 16429; 65app; English.

XX The present invention relates to single exon nucleic acid probes (SENP).

CC CC The present sequence is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful

CC CC for antenatal diagnosis of human genetic disorders

XX SQ Sequence 232 BP; 12 A; 94 C; 5 G; 119 T; 0 U; 2 Other;

Query Match 2.8%; Score 31; DB 4; Length 232;
Best Local Similarity 100.0%; Pred.No. 0.00011;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 975 TTCTCCTCCCCTCTCTCTCTCTCTCTCTCTCTCACG 1005
|-----|
DB 91 TTCTCCTCCCCTCTCTCTCTCTCTCTCTCTCTCACG 121
|-----|

RESULT 15

ABA49623

ID ABA49623 standard; DNA; 232 BP.

XX AC ABA49623;

XX DT 01-FEB-2002 (first entry)

XX DE Human breast cell single exon nucleic acid probe #8318.

XX XX Human; microarray; single exon probe; gene expression; breast; disease;

KW cancer; ss.

XX OS Homo sapiens.

XX PN WO200157271-A2.

PD O9-AUG-2001.

XX 30-JAN-2001; 2001WO-US000662.

XX 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00609408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SRP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.

DR New spatially-addressable set of single exon nucleic acid probes, useful

PT PT for measuring gene expression in sample derived from human breast,

PT comprises number of single exon nucleic acid probes.

XX Claim 4; SEQ ID NO 8318; 327pp + Sequence Listing; English.

XX The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived

CC from human breast and BT 474 cells. The method involves contacting the

CC probes with a collection of detectably labelled nucleic acids derived

CC from mRNA of human breast, and then measuring the label bound to each

CC probe of the microarray. The probes are useful for verifying the

CC expression of regions of genomic DNA predicted to encode proteins. They

CC are useful for gene discovery, and for determining predisposition and/or

CC prognosing breast disease. Gene expression analysis is useful for

CC assessing the toxicity of chemical agents on cells. The microarray of

CC this invention presents a far greater diversity of probes for measuring

CC gene expression, with far less bias than expressed sequence tag

CC microarrays. The method is suitable for rapid production of functional

Search completed: February 21, 2004, 21:44:41
Job time : 487 secs

Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
TREATING ATHEROSCLEROSIS
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/979,608A
FILING DATE: 26-NOV-1997
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/048,547
FILING DATE: 03-JUN-1997
APPLICATION NUMBER: US 60/031,930
FILING DATE: 27-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1362 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...696
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-08-979-608A-12
Query Match 2.6%; Score 29; DB 4; Length 1362;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 972 GTCTTCCTCCCTCCTCCTCCTCCTCCT 1000
Db 84 GTCTTCCTCCCTCCTCCTCCTCCTCCT 56
RESULT 6
US-09-517-849-12/c
Sequence 12, Application US/09517849
Patent No. 6605588
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
Lees, Robert S.
Law, Simon W.
Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
TREATING ATHEROSCLEROSIS
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/517,849
FILING DATE: 02-Mar-2000
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/979,608
FILING DATE: 26-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1362 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...696
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-517-849-12
Query Match 2.6%; Score 29; DB 4; Length 1362;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 972 GTCTTCCTCCCTCCTCCTCCTCCTCCT 1000
Db 84 GTCTTCCTCCCTCCTCCTCCTCCTCCT 56
RESULT 7
US-09-616-289-12/c
Sequence 12, Application US/09616289
Patent No. 6632923
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
APPLICANT: Lees, Robert S.
APPLICANT: Law, Simon W.
APPLICANT: Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
ATHEROSCLEROSIS
FILE REFERENCE: 10797-004001
CURRENT APPLICATION NUMBER: US/09/616,289
CURRENT FILING DATE: 2000-07-14
PRIORITY APPLICATION NUMBER: US 09/517,849
PRIORITY FILING DATE: 2000-03-02
PRIORITY APPLICATION NUMBER: US 08/979,608
PRIORITY FILING DATE: 1997-11-26
PRIORITY APPLICATION NUMBER: US 60/031,930
PRIORITY FILING DATE: 1996-11-27
PRIORITY APPLICATION NUMBER: US 60/048,547
PRIORITY FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 1362
TYPE: DNA
ORGANISM: Oryctolagus cuniculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(696)
US-09-616-289-12

```

: FILE REFERENCE: 10797-004001
:
: CURRENT APPLICATION NUMBER: US/09/616,289
:
: CURRENT FILING DATE: 2000-07-14
:
: PRIOR APPLICATION NUMBER: US 09/517,849
:
: PRIOR FILING DATE: 2000-03-02

```

; CURRENT FILING DATE: 2000-07-14
 ; PRIOR APPLICATION NUMBER: US 09/517,849
 ; PRIOR FILING DATE: 2000-03-02

;; PRIOR APPLICATION NUMBER: US 08/979,608
;; PRIOR FILING DATE: 1997-11-26
;; PRIOR APPLICATION NUMBER: US 60/031,930
;; PRIOR FILING DATE: 1996-11-27
;; PRIOR APPLICATION NUMBER: US 60/048,547
;; PRIOR FILING DATE: 1997-06-03
;; NUMBER OF SEQ ID NOS: 53
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 13
;; LENGTH: 1422
;; TYPE: DNA
;; ORGANISM: Oryctolagus cuniculus
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(756)
US-09-616-289-13

Query Match 2.6%; Score 29; DB 4; Length 1422;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 1000
Db 144 GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 116

RESULT 11
US-08-979-608A-11/c
; Sequence 11, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-No. 6355451-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1617 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:

US-08-979-608A-11

Query Match 2.6%; Score 29; DB 4; Length 1617;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 1000
Db 339 GTCCTCCTCCTCCTCCTCCTCCTCCTCCT 311

RESULT 12
US-09-517-849-11/c
; Sequence 11, Application US/09517849
; Patent No. 6605588
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/517,849
; FILING DATE: 02-Mar-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/979,608
; FILING DATE: 26-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1617 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-517-849-11

Query Match 2.6%; Score 29; DB 4; Length 1617;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 972 GTCCTCCTCCTCCTCCTCCTCCTCCTCCT 1000
Db 339 GTCCTCCTCCTCCTCCTCCTCCTCCTCCT 311

RESULT 13
US-09-616-289-11/c
; Sequence 11, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.

```

Query Match      2.6%; Score 29; DB 1; Length 3728;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      973 TCTTCCTCCTCCTCCTCCTCCTCCTCCTC 1001
          |||||
Db       3385 TCTTCCTCCTCCTCCTCCTCCTCCTCCTC 3357

```

Search completed: February 21, 2004, 23:47:00
Job time : 115 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 21:36:35 ; Search time 468 Seconds
(without alignments)
8260.289 Million cell updates/sec

Title: US-10-041-615-33

Perfect score: 1104
Sequence: 1 atgataaaactaggcccttacttcacccacaaaataaaactaaagattaaagat 60

Scoring table: OLRGP NUC
Gapop 60.0 , Gapext 60.0

Searched: 2308684 seqs, 1750822206 residues

Word size : 10

Total number of hits satisfying chosen parameters: 1853763

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:**

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:**
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:**
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:**
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:**
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:**
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:**
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:**
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:**
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:**
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:**
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:**
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:**
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:**
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:**
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:**
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:**
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:**
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1104	100.0	1104	15	US-10-041-615-33
2	461	41.8	4319	14	US-10-017-161-1619
3	461	41.8	4319	15	US-10-292-798-1293
4	409	37.0	1526	9	US-09-728-952-40
5	176	15.9	536	10	US-09-791-932-35
C 6	159	14.4	556	11	US-09-801-944B-64
C 7	103	9.3	594	9	US-09-812-102-90
C 8	31	2.8	232	9	US-09-864-761-19928
9	31	2.8	384	9	US-09-864-761-3149
C 10	31	2.8	534	15	US-10-260-238-5532
C 11	31	2.8	2000	15	US-10-260-238-1941
C 12	31	2.8	2764	9	US-09-893-666A-1
C 13	30	2.7	123	9	US-09-864-761-27614
14	30	2.7	180	14	US-10-029-386-17166
15	30	2.7	207	9	US-09-864-761-22442

Sequence 17651, A
Sequence 1021, Ap
Sequence 2236, Ap
Sequence 10977, A
Sequence 871, App
Sequence 5679, Ap
Sequence 6520, Ap
Sequence 20250, A
Sequence 3466, Ap
Sequence 273400, A
Sequence 20317, A
Sequence 113575, A
Sequence 264, App
Sequence 24, Appl
Sequence 23, Appl
Sequence 665, App
Sequence 25, Appl
Sequence 26, Appl
Sequence 25, Appl
Sequence 27, Appl
Sequence 4415, A
Sequence 399, App
Sequence 400, App
Sequence 401, App
Sequence 11, Appl
Sequence 497, App
Sequence 496, App
Sequence 496, App
Sequence 1377, Ap

276 9 US-09-864-761-17651
331 9 US-09-728-446-1021
393 10 US-09-803-719-2236
449 9 US-09-864-761-10977
464 9 US-09-864-761-871
471 9 US-09-864-761-5679
504 14 US-10-029-386-6520
558 14 US-10-029-386-20250
598 14 US-10-029-386-3466
661 15 US-10-027-632-273400
736 14 US-10-029-386-20317
739 12 US-10-424-599-113576
2056 14 US-10-007-926A-264
2068 15 US-10-291-172-24
2079 15 US-10-291-172-23
2095 13 US-10-044-090-665
2098 15 US-10-291-172-25
2120 14 US-10-119-428-26
2129 14 US-10-119-428-25
2150 14 US-10-119-428-27
2211 12 US-10-424-599-4415
2317 15 US-10-291-172-399
2317 15 US-10-291-172-400
2317 15 US-10-291-172-401
3953 14 US-10-232-561-11
4450 9 US-09-764-887-497
4450 14 US-10-073-961-497
4454 9 US-09-764-887-496
4454 14 US-10-073-961-496
16033 14 US-10-311-455-1377

ALIGNMENTS

RESULT 1

US-10-041-615-33
; Sequence 33, Application US/10041615
; Publication No. US20040014038A1
; GENERAL INFORMATION:
; APPLICANT: Casman, Stacie J
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Ellerman, Karen
; APPLICANT: Smithson, Glennda
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: No. US20040014038A1el GPCR-Like Proteins and Nucleic Acids Encod
; FILE REFERENCE: 21402-333-061
; CURRENT APPLICATION NUMBER: US/10/041,615
; CURRENT FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: 60/259,552
; PRIOR FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: 60/260,544
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/277,405
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 33
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1080)
US-10-041-615-33

Query Match 100.0%; Score 1104; DB 15; Length 1104;

Best Local Similarity 100.0%; Pred No. 0;

Matches 1104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGATAAACTAGGCCCTTACTTCACCCACAAAATAAACTAAAGATTAAAGAT 60

QY


```
QY 517 CCGCTGTGGGCAAGCTGGCGCTGCATGGAGTACAGCAGCATGAGTCAAGTCCCTCGGGCTG 576
Db 3359 CCGCTGTGGGCAAGCTGGCGCTGCATGGAGTACAGCAGCATGAGTCAAGTCCCTCGGGCTG 3418
QY 577 CCGCTGTGGGCAAGCTGGCGCTGCATGGAGTACAGCAGCATGAGTCAAGTCCCTCGGGCTG 636
Db 3419 CCGCTGTGGGCAAGCTGGCGCTGCATGGAGTACAGCAGCATGAGTCAAGTCCCTCGGGCTG 3478
QY 637 TCCTGCTATATGAAGATCAACCTGGAAGCTGTGCAGCACAGCT 678
Db 3479 TCCTGCTATATGAAGATCAACCTGGAAGCTGTGCAGCACAGCT 3520

RESULT 3
US-10-292-798-1293
; Sequence 1293, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1293
; LENGTH: 4319
; TYPE: DNA
; ORGANISM: Homo sapiens
; LOCATION: source
; FEATURE:
; FEATURE:
; LOCATION: (1)..(4319)
; NAME/KEY: CDS
; LOCATION: (201)..(308)
; NAME/KEY: CDS
; LOCATION: (3128)..(3531)
; NAME/KEY: CDS
; LOCATION: (3660)..(4119)
; NAME/KEY: modified_base
; LOCATION: (1)..(48)
; OTHER INFORMATION: a, t, c, g, unknown or other
US-10-292-798-1293

Query Match 41.8%; Score 461; DB 15; Length 4319;
Best Local Similarity 99.8%; Pred. No. 2.3e-225;
Matches 581; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 97 CGCTCTCTGTTCTACAGAGCCCTCTGCTCTAGTGCCTGGGAAACATCCCTTGCCCTT 156
Db 2940 CGCTCTCTGTTCTACAGAGCCCTCTGCTCTAGTGCCTGGGAAACATCCCTTGCCCTT 2999
QY 157 TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTG 216
Db 3000 TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTG 3059
QY 217 GCTGTGCTGACCTGCTTACCGTGGCCCTTACGGGAAGGGTGGTGTATGTGCTG 276
Db 3060 GCTGTGCTGACCTGCTTACCGTGGCCCTTACGGGAAGGGTGGTGTATGTGCTG 3119
QY 277 GCTGTGCTGACCTGCTTACCGTGGCCCTTACGGGAAGGGTGGTGTATGTGCTG 336
Db 336 GCTGTGCTGACCTGCTTACCGTGGCCCTTACGGGAAGGGTGGTGTATGTGCTG 601
```

```
Db 3120 GGTCTCAAGCTGGCCCTTTTCGGCAAGGGGCTCTGTCAGAGGTGACGGCGTTTGTGCTTACACC 3179
QY 337 GACACCTACGGGGGGGCTTACTCTCATGGCCCTGTGTGAGCGTGGACCAATTACCCAGCTGTG 396
Db 3180 GACACCTAC-GGGGGGCTTACTCTCATGGCCCTGTGTGAGCGTGGACCAATTACCCAGCTGTG 3238
QY 397 GTCTGTGCCCACTGGGGGCCCGGCGCTCCGACAGGCTGSCCGCGCCAGGCTGGTCTGCGCTG 456
Db 3239 GTCTGTGCCCACTGGGGGCCCGGCGCTCCGACAGGCTGSCCGCGCCAGGCTGGTCTGCGCTG 3298
QY 457 GCATCTGAGCCCTTGGTGTCTGTGTCAGACGATGCCCTTGTCTTGTATGCCCATGACCAAG 516
Db 3299 GCATCTGAGCCCTTGGTGTCTGTGTCAGACGATGCCCTTGTCTTGTATGCCCATGACCAAG 3358
QY 517 CCGCTGTGGGCAAGCTGGCCCTGCATGGAGTACAGCAGCATGAGTCAAGTCCCTCGGGCTG 576
Db 3359 CCGCTGTGGGCAAGCTGGCCCTGCATGGAGTACAGCAGCATGAGTCAAGTCCCTCGGGCTG 3418
QY 577 CCGCTCATGCTCTGCTGGTGGCCCTTTGCCATTGGCTTCTGTGGCCAGTGGGATCATCTG 636
Db 3419 CCGCTCATGCTCTGCTGGTGGCCCTTTGCCATTGGCTTCTGTGGCCAGTGGGATCATCTG 3478
QY 637 TCCTGCTATATGAAGATCAACCTGGAAGCTGTGCAGCACAGCT 678
Db 3479 TCCTGCTATATGAAGATCAACCTGGAAGCTGTGCAGCACAGCT 3520

RESULT 4
US-09-728-952-40
; Sequence 40, Application US/09728952
; Patent No. US20020111302A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Kyle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Yamazaki, Vicki
; APPLICANT: Ujwal, Manusha L.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020111302A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; FILE REFERENCE: 799
; CURRENT APPLICATION NUMBER: US/09/728,952
; CURRENT FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: pt_FL_Genes Version 2.0
; SEQ ID NO 40
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (684)..(1016)
; OTHER INFORMATION: a, t, c, g, unknown or other
US-09-728-952-40

Query Match 37.0%; Score 409; DB 9; Length 1526;
Best Local Similarity 99.7%; Pred. No. 9.8e-199;
Matches 579; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 98 GGTCTCTGTTCTACAGAGCCCTCTGCTCTAGTGCCTGGGAAACATCCCTTGCCCTTT 157
Db 422 GGTCTCTGTTCTACAGAGCCCTCTGCTCTAGTGCCTGGGAAACATCCCTTGCCCTTT 481
QY 158 GCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTGG 217
Db 482 GCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTGG 541
QY 218 GTGTGCTGACCTGCTTTCACCGTGGCCCTTACCGGAAGGGTGGTGTATGTGCTG 277
Db 542 GTGTGCTGACCTGCTTTCACCGTGGCCCTTACCGGAAGGGTGGTGTATGTGCTG 601
```



```

Db      327  TTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCACC 357

RESULT 10
US-10-260-238-5532/c
; Sequence 5532, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 5532
; LENGTH: 534
; TYPE: DNA
; ORGANISM: Zea mays
US-10-260-238-5532

Query Match      2.8%; Score 31; DB 15; Length 534;
Best Local Similarity 100.0%; Pred. No. 2,9e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0;

Qy      972  GTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 1002
Db      285  GTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 255

RESULT 11
US-10-260-238-1941/c
; Sequence 1941, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 1941

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 21:26:30 ; Search time 2795 Seconds
(without alignments)

11795.294 Million cell updates/sec

Title: US-10-041-615-33

Perfect score: 1104

Sequence: 1 atgataaaactaggccctta.....gtgtcatttgctttatgact 1104

Scoring table:

OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 10

Total number of hits satisfying chosen parameters: 25639868

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estlin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_estc.*

9: gb_estli.*

10: gb_est2.*

11: gb_hic.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_man.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	103	9.3	594	9 AI659965	AI659965 we64e01.x
C 2	34	3.1	143	13 BY346000	BY346000 BY346000
C 3	32	2.9	203	28 BH689531	BH689531 BOMH51TR
C 4	32	2.9	240	10 AW532693	AW532693 UI-R-B50-

C	5	32	2.9	271	28	BH542697
	6	32	2.9	506	14	CD845470
	7	32	2.9	507	14	CD835500
	8	32	2.9	539	14	CD840534
C	9	32	2.9	570	28	BZ499662
C	10	32	2.9	613	28	BH667854
C	11	32	2.9	628	28	BH661526
C	12	32	2.9	646	28	BZ430821
C	13	32	2.9	701	28	BH977513
C	14	32	2.9	702	28	BH938360
C	15	32	2.9	706	14	CD874215
C	16	32	2.9	729	28	BH537878
C	17	32	2.9	769	28	BZ453671
C	18	32	2.9	786	28	BZ068671
C	19	32	2.9	812	28	BH704785
C	20	32	2.9	836	28	BZ504153
C	21	32	2.9	889	29	CNS0403A
C	22	32	2.9	1121	14	CK209771
C	23	31	2.8	185	14	CD942005
C	24	31	2.8	210	29	CG026240
C	25	31	2.8	328	28	BH061458
C	26	31	2.8	364	14	CF063130
C	27	31	2.8	515	28	AZ039092
C	28	31	2.8	552	28	AZ039077
C	29	31	2.8	570	14	CB334566
C	30	31	2.8	638	29	CG441247
C	31	31	2.8	661	14	CB079907
C	32	31	2.8	664	14	CA235347
C	33	31	2.8	686	14	CA195996
C	34	31	2.8	767	29	AB055167
C	35	31	2.8	785	14	CB668345
C	36	31	2.8	791	29	CNS04PSZ
C	37	31	2.8	819	29	CG047970
C	38	31	2.8	833	29	CG253509
C	39	31	2.8	880	29	CG132477
C	40	31	2.8	915	29	CNS02Z1L
C	41	31	2.8	927	29	CNS03Y04
C	42	31	2.8	951	29	CNS02S29
C	43	31	2.8	953	29	CNS02UFR
C	44	31	2.8	1177	14	CK161859
C	45	30	2.7	94	28	AZ804535

ALIGNMENTS

RESULT 1
AI659965/c
LOCUS
DEFINITION
IMAGE:2345880 3' similar to SW:EBI2 HUMAN P32249 EBV-INDUCED G
PROTEIN-COUPLED RECEPTOR 2 ; mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AI659965 594 bp mRNA linear EST 18-DEC-1999

we64e01.x1 Soares thymus NHFT Homo sapiens cDNA clone

IMAGE:2345880 3' similar to SW:EBI2 HUMAN P32249 EBV-INDUCED G

PROTEIN-COUPLED RECEPTOR 2 ; mRNA sequence.

AI659965
AI659965.1 GI:4763535

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 594)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapsb-remail.nih.gov

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (infoimage.lnl.gov) for further information.

Insert Length: 1014 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 408.

Location/Qualifiers

1..594

source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2345880"
/dev_stage="fetal"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares thymus NHFTh"
/note="Organ: thymus, pooled; Vector: pTT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I,
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(GT) primer [5'.
TGTACCAATCTGAAGTGGGACGGCGCGCAAGTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaudo. "

ORIGIN

      Query Match
      Best Local Similarity 100.0%; Score 103; DB 9; Length 594;
      Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 CTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTTGTCTTACACCGACCTACGGG 349
Db 530 CTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTTGTCTTACACCGACCTACGGG 471
QY 350 GGGTCTACTCATGGCTGTGTGAGCGTGAACCATACCAGC 392
Db 470 GGGTCTACTCATGGCTGTGTGAGCGTGAACCATACCAGC 428

RESULT 2
BY346000
LOCUS BY346000 143 bp mRNA linear EST 12-DEC-2002
DEFINITION BY346000 RIKEN full-length enriched, whole joints Mus musculus cDNA
clone L730017L22 5', mRNA sequence.
ACCESSION BY346000.1 GI:26575488
VERSION
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 143)
AUTHORS
Nikaido, I., Oatso, N., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojochori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Rume, D.A., Quackenbush, J., Schram, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
Chochia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Giesi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawai, J.H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Seiple, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, N.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlesch, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wyman-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Inotani, K., Ishii, Y.,
Itoh, M., Kigawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.
ANALYSIS OF THE MOUSE TRANSCRIPTOME BASED ON FUNCTIONAL ANNOTATION

```



```

972 GTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 1003
|||||
132 GTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 101

RESULT 6
CD845470          506 bp  mRNA  linear  EST 11-JUL-2003
LOCUS
DEFINITION
RFO2.143F14F011228 RFO2 Brassica napus cDNA clone RFO2143F14, mRNA
sequence.
ACCESSION
CD845470
CD845470.1 GI:32529292
EST.
Brassica napus (rape)
ORGANISM
Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 506)
REFERENCE
AUTHORS
Genoplante.
TITLE
Brassica napus
JOURNAL
Unpublished (2003)
COMMENT
Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.infobiogen.fr).
FEATURES
source
1..506
/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="samourai (restored line)"
/db_xref="taxon:3708"
/clone="RFO2143F14"
/tissue_type="anthers"
/clone_lib="RFO2"

Query Match 2.9%; Score 32; DB 14; Length 506;
Best Local Similarity 100.0%; Pred. No. 9.6e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
973 TCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCAC 1004
|||||
86 TCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCAC 117

RESULT 7
CD839500          507 bp  mRNA  linear  EST 10-JUL-2003
LOCUS
DEFINITION
RFO2.115J02F010530 RFO2 Brassica napus cDNA clone RFO2115J02, mRNA
sequence.
ACCESSION
CD839500
CD839500.1 GI:32521440
EST.
Brassica napus (rape)
ORGANISM
Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 507)
REFERENCE
AUTHORS
Genoplante.
TITLE
Brassica napus
JOURNAL
Unpublished (2003)
COMMENT
Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10

This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.infobiogen.fr).
FEATURES
source
1..539
/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="samourai (restored line)"
/db_xref="taxon:3708"
/clone="RFO2115J18"
/tissue_type="anthers"
/clone_lib="RFO2"

Query Match 2.9%; Score 32; DB 14; Length 539;
Best Local Similarity 100.0%; Pred. No. 9.7e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
973 TTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCAC 1004
|||||
119 TTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCAC 150

RESULT 8
CD840534          539 bp  mRNA  linear  EST 10-JUL-2003
LOCUS
DEFINITION
RFO2.119L18F010605 RFO2 Brassica napus cDNA clone RFO2119L18, mRNA
sequence.
ACCESSION
CD840534
CD840534.1 GI:32522474
EST.
Brassica napus (rape)
ORGANISM
Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 539)
REFERENCE
AUTHORS
Genoplante.
TITLE
Brassica napus
JOURNAL
Unpublished (2003)
COMMENT
Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.infobiogen.fr).
FEATURES
source
1..539
/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="samourai (restored line)"
/db_xref="taxon:3708"
/clone="RFO2119L18"
/tissue_type="anthers"
/clone_lib="RFO2"

Query Match 2.9%; Score 32; DB 14; Length 539;
Best Local Similarity 100.0%; Pred. No. 9.7e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
973 TTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCAC 1004
|||||
119 TTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCAC 150

RESULT 9
BZ499662/c
LOCUS
DEFINITION
BZ499662
BONLU34TR BO 1.6.2 KB tot Brassica oleracea genomic clone BONLU34,
genomic survey sequence.
570 bp  DNA  linear  GSS 16-DEC-2002
```

```

This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.infobiogen.fr).
FEATURES
source
1..507
/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="samourai (restored line)"
/db_xref="taxon:3708"
/clone="RFO2115J02"
/tissue_type="anthers"
/clone_lib="RFO2"

ORIGIN
Query Match 2.9%; Score 32; DB 14; Length 507;
Best Local Similarity 100.0%; Pred. No. 9.6e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
973 TCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCAC 1004
|||||
87 TCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCAC 118

RESULT 8
CD840534          539 bp  mRNA  linear  EST 10-JUL-2003
LOCUS
DEFINITION
RFO2.119L18F010605 RFO2 Brassica napus cDNA clone RFO2119L18, mRNA
sequence.
ACCESSION
CD840534
CD840534.1 GI:32522474
EST.
Brassica napus (rape)
ORGANISM
Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 539)
REFERENCE
AUTHORS
Genoplante.
TITLE
Brassica napus
JOURNAL
Unpublished (2003)
COMMENT
Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.infobiogen.fr).
FEATURES
source
1..539
/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="samourai (restored line)"
/db_xref="taxon:3708"
/clone="RFO2119L18"
/tissue_type="anthers"
/clone_lib="RFO2"

Query Match 2.9%; Score 32; DB 14; Length 539;
Best Local Similarity 100.0%; Pred. No. 9.7e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
973 TTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCAC 1004
|||||
119 TTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCAC 150

RESULT 9
BZ499662/c
LOCUS
DEFINITION
BZ499662
BONLU34TR BO 1.6.2 KB tot Brassica oleracea genomic clone BONLU34,
genomic survey sequence.
570 bp  DNA  linear  GSS 16-DEC-2002
```



```

ACCESSION      CD874215
VERSION        CD874215.1  GI:32558031
KEYWORDS       EST.
SOURCE         Triticum aestivum (bread wheat)
ORGANISM       Triticum aestivum
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
               Pooideae; Triticeae; Triticeae; Triticum.
REFERENCE      1. (bases 1 to 706)
               Genoplatane.
               Genoplatane, a major partnership french program in plant genomics
               Unpublished (2003)
               Contact: Genoplatane
               Genoplatane
               93, rue Henri Rochefort 91025 EVRY CEDEX France
               Tel: 33 1 69 47 54 00
               Fax: 33 1 69 47 54 10
               This sequence has been generated in the framework of the french
               plant genomics programme 'Genoplatane' (http://www.genoplatane.com
               and http://genoplatane-info.infobiogen.fr).
FEATURES       1. .706
               source
               /organism="Triticum aestivum"
               /mol_type="mRNA"
               /cultivar="recital"
               /db_xref="taxon:4565"
               /clone="AZ09101K13"
               /tissue_type="leaf"
               /clone_lib="AZ03"

ORIGIN
Query Match      2.9%; Score 32; DB 14; Length 706;
Best Local Similarity 100.0%; Pred.No. 0.0001;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      974  CTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCACC 1005
Db      95  CTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCACC 64

Search completed: February 21, 2004, 23:44:56
Job time : 2815 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 19:17:05 ; Search time 43 Seconds
(without alignments)
1328.874 Million cell updates/sec

Title: US-10-041-615-34

Perfect score: 1876

Sequence: 1 MIKLGYPFTPTKTKIKND.....SETPSITQAGSMFLAEHW 360

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1876	100.0	360	23	ABG32286 Human GPCR16 prote
2	1505	80.3	377	22	ABG14682 Novel human diago
3	632.5	33.7	361	15	AAK54080 Epstein Barr virus
4	632.5	33.7	361	19	AAW53623 Epstein Barr virus
5	632.5	33.7	361	21	AAAY90630 Human G protein-co
6	632.5	33.7	361	24	ABP81814 Human EBV-induced
7	630.5	33.6	361	21	AAAY90664 Human mutant G pro
8	555	29.6	110	23	ABG66698 Human novel polytpe
9	417	22.2	177	22	AAU25588 Human G Protein-Co

10	395.5	21.1	339	21	AAAY90652 Human mutant G pro
11	391.5	20.9	339	15	AAK53752 Seven transmembran
12	391.5	20.9	339	18	AAW07617 Human G-protein th
13	391.5	20.9	339	19	AAW48733 Human R12 seven tr
14	391.5	20.9	339	21	AAE21697 Human 7TM receptor
15	391.5	20.9	339	21	AAAY90618 Human G protein-co
16	391.5	20.9	339	22	AAH82853 Human P2Y-like GPC
17	391.5	20.9	339	23	AAU98065 Human cysteinyl le
18	391.5	20.9	339	23	AAU91239 Human 7 transmembr
19	391.5	20.9	339	24	ABU11906 Human G-protein co
20	391.5	20.9	367	24	ABP81909 Human G-protein-co
21	377.5	20.1	177	22	AAU19222 Human PAR1 type th
22	355	18.9	371	23	ABG35299 Human PAR1 type th
23	352.5	18.8	402	23	ABG35298 Human PAR1 type th
24	352.5	18.8	425	23	ABG35300 Fragment of the hu
25	340	18.1	425	15	AAK60698 Human thrombin rec
26	339.5	18.1	425	13	AAK27240 Human thrombin rec
27	339.5	18.1	425	19	AAW51407 Human protease-act
28	339.5	18.1	425	20	AAK49570 Human thrombin rec
29	339.5	18.1	425	23	ABG80697 Human thrombin pro
30	339.5	18.1	425	23	AAE17032 Human thrombin pro
31	339.5	18.1	425	24	ABR47449 Breast cancer asso
32	339.5	18.1	425	24	ABP81919 Human thrombin rec
33	339.5	18.1	425	24	ABG73511 Human thrombin pro
34	339.5	18.1	426	21	AAK45035 Human thrombin rec
35	339.5	18.1	892	18	AAW16314 X. laevis thrombin
36	332.5	17.7	420	24	ABG73510 Human P2Y purinoco
37	332	17.7	537	23	AAU74538 Human placenta G-p
38	331	17.6	328	17	AAK91225 Human P2-purinergi
39	329	17.5	328	22	AAE04393 Human purinergic r
40	329	17.5	328	24	ABP81869 Non-endogenous hum
41	328	17.5	355	22	ABK56343 Human placenta pur
42	324	17.3	328	18	AAW09433 Human G-protein re
43	324	17.3	355	18	AAW07618 Human G-protein ch
44	324	17.3	355	19	AAW49807 Human CC chemokine
45	324	17.3	355	20	AAW97868

ALIGNMENTS

RESULT 1
ABG32286

ID ABG32286 standard; Protein; 360 AA.

AC ABG32286;

DT 15-NOV-2002 (first entry)

DE Human GPCR16 protein.

XX G protein coupled receptor related protein; human; GPCR;

KW cardiomyopathy; atherosclerosis; diabetes; cancer; stroke;

KW Von Hippel-Lindau syndrome; Alzheimer's disease; Huntington's disease;

KW hypercalcaemia; Parkinson's disease; Huntington's disease;

KW cerebral palsy; epilepsy; Lesch-Nyhan syndrome; multiple sclerosis;

KW ataxia-telangiectasia; leukodystrophy; addiction; anxiety; depression;

KW pain; obesity; Crohn's disease; osteoporosis; haemophilia; asthma;

KW inflammatory bowel disease; infertility; hypertension; scleroderma;

KW arthritis; human immunodeficiency virus; autoimmune disease; HIV;

XX infection; graft-versus-host disease.

XX Homo sapiens.

XX WO200264793-A2.

PD 22-AUG-2002.

PF 03-JAN-2002; 2002MO-US000056.

PR 03-JAN-2001; 2001US-259552P.

PR 09-JAN-2001; 2001US-260544P.

PR 20-MAR-2001; 2001US-277405P.

Db 10 LFTYALLVPSALGNILALCLTCQKRNKCTGIYLVHLAVSDLLFTVALPGRVVCYVLGS 69
 QY 95 SWPFGKGLCRLTAFVLYTDTGGVYLMACVSDHYPAVVCVCAHWGPRLTAGRLVCVAI 154
 Db 70 SWPFGKGLCRLTAFVLYTDTGGVYLMACVSDHYPTVVCVCAHWGPRLTAGRLVCVAI 129
 QY 155 WTLVLQTMPLLMPTKPLVGLKACMEYSWESVGLPLMLVAFVAFGCGPVGIILSC 214
 Db 130 WTLVLQTMPLLMPTKPLVGLKACMEYSWESVGLPLMLVAFVAFGCGPVGIILSC 189
 QY 215 YNKITWKLCS ----- 224
 Db 190 YNKITWKLCSAGRTQXPAGKDTTGGASFGGSPDQPEKTPPAGAAQEDPVTTPRTTPAG 249
 QY 225 -----TAREDPVTSRKGHRWGCLLTLLMLVAVVVCFSPYHLNIKOFMA 268
 Db 250 QPERTQXPAGKDATGCTAREDPVTSRKGHRWGCLLTLLMLVAVVVCFSPYHLNIKOFMA 309
 QY 269 RGMHLPSCAERRAFLLSLQATVALMNNVCGITPIIYFFASTHYRKMILLGLIKLKGSSSS 328
 Db 310 RGMHLPSCAERRAFLLSLQATVALMNNVCGIDPIIYFFASTHYRKMILLGLIKLKGSSSS 369
 QY 329 SSSSSS 334
 Db 370 SSSSST 375

RESULT 3
 AAR54080
 ID AAR54080 standard; Protein; 361 AA.
 XX AC AAR54080;
 XX DT 25-MAR-2003 (updated)
 DT 03-FEB-1995 (first entry)
 XX DE Epstein Barr virus induced (EBI-2) polypeptide.
 XX KW Epstein Barr virus; EBV; induction; detection; diagnosis;
 KW lymphocytes; antigen; growth; differentiation; mediator;
 KW infectious mononucleosis.
 OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT Modified-site 8
 FT Region /note= "Potential N-linked glycosylation site."
 FT Region 32..57 /label= Hydrophobic region.
 FT Region 78..95 /label= Hydrophobic region.
 FT Region 106..127 /label= Hydrophobic region.
 FT Region 126..134 /label= Hydrophobic region.
 FT /note= "This sequences motif
 (S-[I/V]-D-R-[V/F]-X-X-X) is highly
 conserved among a large number of G-protein
 coupled receptors."
 FT Region 150..168 /label= Hydrophobic region.
 FT Region 193..215 /label= Hydrophobic region.
 FT Region 242..265 /label= Hydrophobic region.
 FT Region 288..312 /label= Hydrophobic region.
 FT WO9412519-A1.
 PN XX
 PD 09-JUN-1994.
 XX
 PF 08-OCT-1993; 93WO-US09636.
 XX

PR 25-NOV-1992; 92US-0980518.
 XX (BCHM) BRIGHAM & WOMENS HOSPITAL.
 PA Birkenbach M, Kieff E;
 PI N-PSDB; AAQ64126.
 XX WPI; 1994-200183/24.
 DR DNA coding for Epstein Barr Virus induced (EBI) polypeptide(s)
 XX and antibodies to EBI1, 2 and 3 - useful for detecting EBV by
 PT hybridisation or by immunoassay
 PT Claim 8; Page 58-60; 84pp; English.
 PS EBV infected B lymphocytes recapitulate features of antigen
 CC stimulation in enlarging, increasing RNA synthesis, expressing
 CC activation antigens and adhesion molecules, secreting Ig and
 CC proliferating. Unlike antigen stimulated B lymphocytes, EBV
 CC infected B lymphocytes continue to proliferate (in vitro) as
 CC immortalised lymphoblastoid cell lines. Because of the similar
 CC effects of EBV and antigen, EBV induced genes are likely to include
 CC mediators of antigen induced B lymphocyte growth or differentiation.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC SQ Sequence 361 AA;
 Query Match 33.7%; Score 632.5; DB 15; Length 361;
 Best Local Similarity 37.8%; Pred. No. 1.1e-58;
 Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;
 QY 2 IKLGPYTPPTKI-KTKIKDLNVESTVKIFLESLEYTALLVPSALGNILALCLTCQSR 60
 Db 3 IQMANNFTPPSATPQGNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVIVQNRK 61
 QY 61 KINCTGIYLVHLAVSDLLFTVALPGRVVCVGLSSWPFKGKLCRLTAFVLYTDTGGVYL 120
 Db 62 KINSTLYSTNLVISDILFTTALPTPIAYANGFDWRIGDALCRITAFYINTYAGVNF 121
 QY 121 MACVSDHYPAVVCVCAHWGPRLTAGRLVCVAIWTVLQTMPLLMPTKPLVGLKAC 180
 Db 122 MTCLSIDRFTAVVHPLRYNKRKRIEHAQKGVCFWILVFAQTLLPLINPMSKOEAEITC 181
 QY 181 MEYSSMESVGLPLMLVAFVAFGCGPVGIILSCYMKITWKLCSAREDPVTSRKGHRWR 240
 Db 182 MEYFNFEETKSLPWILLGACFGVYVPLIILICYQICKLFTAKQPLTEKSGVKK 241
 QY 241 GCLLTLLMLVAVVVCFSPYHLNIKQPMARGM--LHLPSCAERRAFLLSLQATVALMNNVC 298
 Db 242 ALNTIILIIWVFLCTPYHVAIIQHMIXKLRFSNFECSQRHSFOISLHFTVCLMNFNC 301
 QY 299 GITPIIYFFASTHYRKMILLGLIKLKGSSSSSSSSSSSTPGKAS--SETPTSITCARGS 352
 Db 302 CMDPFIYFFACKGYKRVKVMKLRQVSVSISSAVKSAPENSREMTETQMHIHKS 358
 RESULT 4
 AAR53623
 ID AAR53623 standard; Protein; 361 AA.
 XX AC AAR53623;
 XX DT 09-JUL-1998 (first entry)
 XX DE Epstein Barr virus induced gene 2 (EBI-2).
 XX KW Assessing; monitoring; foetal development; placental development;
 KW Epstein Barr virus; EBV; induced gene 2; EBI-2.
 OS Homo sapiens.
 PN US5744301-A.
 XX

QY 299 GITPIIYFPAFTHYKWLGIILKLGSSSSSSSTPGKAS---SETPSITQARGS 352
 DB 302 CMDPIIYFPAFTHYKWLGIILKLGSSSSSSSTPGKAS---SETPSITQARGS 358

RESULT 6
 ABP81814
 ID ABP81814 standard; Protein; 361 AA.
 AC ABP81814;
 XX
 DT 04-MAR-2003 (first entry)
 XX
 DE Human EBV-induced gene 2 protein SEQ ID NO:112.
 XX
 KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;
 KW G protein-coupled receptor modulator; antibody; immune-related disease;
 KW growth-related disease; cell regeneration-related disease; AIDS; cancer;
 KW immunological-related cell proliferative disease; autoimmune disease;
 KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
 KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
 KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;
 KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
 KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
 KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
 KW ulcer.
 XX
 OS Homo sapiens.
 XX
 PN WO200261087-A2.
 XX
 XX 08-AUG-2002.
 XX
 XX 19-DEC-2001; 2001WO-US50107.
 XX
 XX 19-DEC-2000; 2000US-257144P.
 XX
 PA (LIFE-) LIFESPAN BIOSCIENCES INC.
 XX
 XX Burner GC, Roush CL, Brown JP;
 XX
 XX WPI; 2003-046718/04.
 XX
 XX N-PSDB; ABZ42660.
 XX
 XX New isolated antigenic peptides e.g., for G protein-coupled receptors
 PT (GPCR), useful for diagnosing and designing drugs for treating
 PT conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,
 PT cancer or autoimmune diseases -
 XX
 XX Disclosure; Fig 1; 523pp; English.

CC The present invention describes antigenic peptides (I) comprising:
 CC (a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino
 CC acids. Also described: (1) an assay for the detection of a particular
 CC G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;
 CC and (2) an isolated antibody having high specificity and high affinity
 CC or avidity for a particular GPCR. (I) can be used as GPCR modulators and
 CC in gene therapy. The antigenic peptides for GPCRs are useful in detecting
 CC an antibody against a particular GPCR, and in the production of specific
 CC antibodies. The peptides and antibodies are also useful for detecting the
 CC presence or absence of corresponding GPCRs. The antigenic peptides for
 CC GPCRs and antibodies are useful for diagnosing and designing drugs for
 CC treating immune-related diseases, growth-related diseases, cell
 CC regeneration-related disease, immunological-related cell proliferative
 CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,
 CC atherosclerosis, bacterial, fungal, protozoan or viral infections,
 CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute
 CC inflammation, allergies, Crohn's disease, diabetes, graft versus host
 CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,
 CC anxiety, depression, schizophrenia, dementia, mental retardation, memory
 CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,
 CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or
 CC any other disorder in which GPCRs are involved. The antibodies may be

CC used in immunoassays and immunodiagnosis. ABZ42523 to ABZ42869 encode
 CC GPCR proteins given in ABP81675 to ABP82018, which are used in the
 CC exemplification of the present invention.
 XX
 SQ Sequence 361 AA;
 Query Match 33.7%; Score 632.5; DB 24; Length 361;
 Best Local Similarity 37.8%; Pred. No. 1.1e-58;
 Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;
 QY 2 IKLGPYFTPTTKI-KTKIKDLNVESTVTKIFLSLEYTALLVFSALGNILALCLTCQKSR 60
 DB 3 IQMANNFTPEATPQGNDCDLAYAHSTARIVM-PLHYSLVFIIGLVNLLALVVIQNRK 61
 QY 61 KINCTGYLVHLAVSDLLFTVLPGRVVCVYLGSSWPFQKGLCRLTAFVLYTDTYGGVYL 120
 DB 62 KINSTLYSTNLVSDILFTLTALPTRIAYYAMGFDNRIGDALCRITALVFIYINTYAGVNF 121
 QY 121 MACVSDHYPAVCAHWGPPRLTAGRLVVCVAITVILLOTMPILLMPKTLVCKLAC 190
 DB 122 MTLCLSIDRFIAVHPLRYNKKIRIEHAKGVICIFVWILVFAQTLLPLINPMKQEAERITC 181
 QY 181 MEYSSMESVLGLPLMVLVAFAGFCGPGVGIILSCYMKITWKLCTAREDDPVSRRGHRWR 240
 DB 182 MEYPNFEETKSLPWILLGACFIVPLIILILICYQICCKLFRITAKQNPLETEKSGWNKK 241
 QY 241 GCLLTILMLVAVVVCSPYHLNKKQFWARGM--LHLPSCAERRAFLLSLQATVALMMNC 298
 DB 242 ALNTIILIIIVFVLCPTFYHVAIOHMKIKLRFSNFELECSQRHSFOISLHFTVCLMNFNC 301
 QY 299 GITPIIYFPAFTHYKWLGIILKLGSSSSSSSTPGKAS---SETPSITQARGS 352
 DB 302 CMDPIIYFPAFTHYKWLGIILKLGSSSSSSSTPGKAS---SETPSITQARGS 358

RESULT 7
 AAY90664
 ID AAY90664 standard; Protein; 361 AA.
 AC AAY90664;
 XX
 DT 21-AUG-2000 (first entry)
 XX
 DE Human mutant G protein-coupled receptor EBI2 (L243K).
 KW G protein-coupled receptor; GPCR; constitutively active;
 KW intracellular loop 3; transmembrane domain 6; drug screening;
 KW agonist; antagonist; mutant; mutein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 XX WO2000022129-A1.
 XX
 XX 20-APR-2000.
 XX
 XX 12-OCT-1999; 99WO-US23938.
 XX
 XX 13-OCT-1998; 98US-0170496.
 XX
 XX (AREN-) ARENA PHARM INC.
 XX
 XX Behan DP, Chalmers DT, Liaw CW;
 XX
 XX WPI; 2000-329165/28.
 XX
 XX N-PSDB; AAA30730.
 XX
 XX Non-endogenous constitutively activated human G protein-coupled
 PT receptors, useful for identifying agonists for use as pharmaceutical
 PT agents -
 XX
 XX Example 2; Page 262-263; 341pp; English.

CC The invention relates to constitutively active, non-endogenous versions
CC of endogenous human orphan G protein-coupled receptors (GPCRs), AAY90643-
CC AAY90677 and AAY90683-190687, and to DNA encoding them (AA30709-A30743
CC and AAA30775-A30779). The mutant proteins of the invention contain a
CC mutation in a portion of the protein comprising intracellular loop 3
CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC is substituted for an endogenous residue in IC3 at a position 16 amino
CC acids N-terminal of an endogenous proline in TM6 to form a sequence
CC X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
CC or Ala, and is preferably Lys. When the endogenous residue at this
CC position is Lys, this residue is replaced by His, Arg or preferably Ala.
CC The 15 amino acid stretch between the substituted amino acid and the Pro
CC may be endogenous, non-endogenous, or a mixture of endogenous and
CC non-endogenous residues. The constitutively active GPCRs are useful for
CC identifying antagonists, agonists and partial agonists for use as
CC pharmaceutical agents. The mutant proteins are also useful in research
CC settings for elucidating the roles of the receptors in normal and
CC diseased conditions. Antagonists for a particular GPCR are useful for
CC treating diseases and disorders associated with that receptor. Because
CC the novel mutant GPCRs are constitutively active, they can be used
CC directly for screening of compounds without the need for endogenous
CC ligands. Sequences AAY90643- AAY90677 and AAY90683-190687 the mutant
CC human GPCRs of the invention.

XX SQ Sequence 361 AA;
Query Match 33.6%; Score 630.5; DB 21; Length 361;
Best Local Similarity 37.8%; Pred. No. 1.8e-58;
Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;
Qy 2 IKLGYPPTPTKI-KTKIKDLNYESQVKIFLSLFTVALLVPSALGNILALCLTCQKR 60
Db 3 IQWANNTPSPATPQGNDCDLYAAHSTARIVM-PLHSLVFIIGLVGNLLALVIVQNEK 61
Qy 61 KINTGILYVHLAVSDLLFTVALPGRVVCVVLGSSWPFKGLRLFAFLVLYTDTYGGVYL 120
Db 62 KINSTLYSTNLVSDILFTTALPTRIAYYANGFDWRIGDALCRITAFVINTYAGVNF 121
Qy 121 MACVSDHYPAVCAHWGRLRAGRLVCVAITWLVLLQTMPLLLMPTKPLVGLKAC 180
Db 122 MTCLSIDRFIAVHPLRYNKIKRIEHAQVCIFWVILVFAQTUPLINPMKQEARITC 181
Qy 181 MEYSSMESVLGLPLMLVLAFAIGFCGPGVGIILSCYMKITWKLCTAREDPVTSRKGRHW 240
Db 182 MEYFNPEETKSLPWLILGACFIGVVLPLIILICYSQICCKLFRKQNLTEKSGVNNK 241
Qy 241 GCLTLLMLVAVVCFSPYHINIKQFARGM--LHLPSCAERRAFILSLCATVALMMNC 298
Db 242 AKNTIILIVVFLCFTPYHVAIIQHMVKLRFNFLECSQRHSFQISLHFTVCLMNFNC 301
Qy 299 GITPIIYFFASTHYRKWLLGLKLGSSSSSSSSSTPGKAS---SETPSITQAGS 352
Db 302 CMDPIYFFACKGYRKVWRMLKEQVSVISSAVKSAPENSREMTETQMIHNSKS 358
RESULT 8
ID ABG66698 standard; Protein; 110 AA.
XX ABG66698;
XX 30-AUG-2002 (first entry)
XX Human novel polypeptide #33.
XX Human; inflammatory condition; shock; sepsis; immune response;
XX cancer; wound healing; central nervous system disease; haematopoiesis;
XX peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
XX myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
XX cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
XX bone degenerative disorder; periodontal disease; reperfusion injury;
XX lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
XX allergic condition; thrombolysis; thrombosis; coagulation disorder;

XX fungal infection.
XX Homo sapiens.
XX WO200244340-A2.
XX 06-JUN-2002.
XX 30-NOV-2001; 2001WO-US47004.
XX 30-NOV-2000; 2000US-0028952.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
XX Yamazaki V, Ujwal ML, Drmanac RT;
XX WPI; 2002-508509/54.
XX N-PSDB; ABK94922.
XX Novel nucleic acids and polypeptides for diagnosis, treatment of
XX inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
XX disorders, cancer and promoting wound healing -
XX Claim 10; Page 600; 672pp; English.
XX The invention relates to human novel polynucleotides and associated
XX polypeptides. The polynucleotides and polypeptides are useful for
XX treating inflammatory conditions such as arthritis, nephritis, Crohn's
XX disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
XX and cancer and for promoting wound healing. The sequences are used to
XX induce the proliferation of neural cells and regeneration of nerve and
XX brain tissue, and are useful for the treatment of central and peripheral
XX nervous system diseases and neuropathies, such as Alzheimer's disease,
XX Parkinson's disease, Huntington's disease and amyotrophic lateral
XX sclerosis. The sequences are involved in chemotactic or chemokinetic
XX activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
XX cell disorders and platelet disorders such as thrombocytopenia.
XX regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
XX growth, tissue repair, healing of burns, incisions, ulcers, treatment of
XX osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
XX disease. The sequences of the invention are also useful for gut
XX protection or regeneration and treatment of lung or liver fibrosis,
XX reperfusion injury in various tissues, immune deficiencies and disorders
XX including severe combined immunodeficiency (SCID), bacterial or fungal
XX infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
XX gravis, allergic conditions such as asthma, thrombolysis or thrombosis
XX and coagulation disorders. Sequences ABG66666-ABG66758 represent human
XX novel polypeptides of the invention.

XX SQ Sequence 110 AA;
Query Match 29.6%; Score 555; DB 23; Length 110;
Best Local Similarity 99.1%; Pred. NO. 4.3e-51;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 121 MACVSDHYPAVCAHWGRLRAGRLVCVAITWLVLLQTMPLLLMPTKPLVGLKAC 180
Db 1 MACVSDHYPAVCAHWGRLRAGRLVCVAITWLVLLQTMPLLLMPTKPLVGLKAC 60
Qy 181 MEYSSMESVLGLPLMLVLAFAIGFCGPGVGIILSCYMKITWKLCTA 226
Db 61 MEYSSMESVLGLPLMLVLAFAIGFCGPGVGIILSCYMKITWKLCTA 106
RESULT 9
AAU25588
ID AAU25588 standard; Protein; 177 AA.
XX AAU25588;
XX AAU25588;
XX 18-DEC-2001 (first entry)
XX

DE Human G Protein-Coupled Receptor (GPCR) polypeptide #35.
XX Human; G-protein coupled receptor; GPCR; mental disorder; schizophrenia;
KW attention deficit disorder; anxiety; depression; bipolar disorder;
KW neurological disorder; Huntington's disease; dementia; obesity; anorexia;
KW metabolic disorder; Parkinson's disease; Tourette's syndrome; thrombosis;
KW type 2 diabetes; cardiovascular disorder; myocardial infarction; cancer;
KW cardiomyopathy; atherosclerosis; human immunodeficiency virus; HIV;
KW viral infection; immunostimulant; neuroleptic; nootropic; tranquilizer;
KW antidepressant; anorectic; gene therapy.
OS Homo sapiens.
XX
XX WO200162797-A2.
XX
XX 30-AUG-2001.
XX
XX 23-FEB-2001; 2001WO-US05676.
XX
XX 23-FEB-2000; 2000US-0184247.
XX 23-FEB-2000; 2000US-0184303.
XX 23-FEB-2000; 2000US-0184304.
XX 23-FEB-2000; 2000US-0184305.
XX 23-FEB-2000; 2000US-0184397.
XX 02-MAR-2000; 2000US-0186457.
XX 03-MAR-2000; 2000US-0186810.
XX 03-MAR-2000; 2000US-0188064.
XX 13-MAR-2000; 2000US-0188880.
XX 03-APR-2000; 2000US-0194344.
XX 23-JUN-2000; 2000US-0213861.
XX 11-JUL-2000; 2000US-0217369.
XX 11-JUL-2000; 2000US-0217370.
XX 14-JUL-2000; 2000US-0218337.
XX 20-JUL-2000; 2000US-0218492.
XX (PHAA) PHARMACIA & UPJOHN CO.
XX
XX Vogeli G, Wood LS, Parodi LA, Lind P;
PI
XX WPI; 2001-570628/64.
XX N-PSDB; AAS42840.
XX
XX New isolated nucleic acid encoding a new G-protein coupled receptor
PT polypeptide for detecting receptor modulators that can treat mental
PT disorders, such as schizophrenia, anxiety, depression, or obesity -
XX
XX Claim 35; Page 83; 279pp; English.
XX
XX Sequences AAU2554-AU25616 represent human G-protein coupled receptor
CC (GPCR) polypeptides of the invention. The proteins and their associated
CC DNA sequences can be used to identify compounds which bind to GPCR
CC polypeptides and in screening for compounds that modulate GPCR activity.
CC By screening a human subject for the presence of mutations in GPCR DNA, a
CC GPCR-related disorder or a genetic predisposition can be diagnosed. The
CC sequences can also be used for treatment and prevention of mental
CC disorders such as schizophrenia, attention deficit disorder, anxiety,
CC depression, dementia and bipolar disorder, neurological disorders such as
CC Huntington's disease, Parkinson's disease and Tourette's syndrome,
CC metabolic disorders such as obesity, anorexia and type 2 diabetes,
CC cardiovascular disorders such as thrombosis, myocardial infarction,
CC cardiomyopathy and atherosclerosis, viral infections caused by HIV and
CC cancers.
XX
XX Sequence 177 AA;
SQ
Query Match 22.2%; Score 417; DB 22; Length 177;
Best Local Similarity 58.9%; Pred. No. 4.3e-36;
Matches 93; Conservative 8; Mismatches 17; Indels 40; Gaps 5;
OY 101 GLCLTAPLVYDTRGGVYLMAVSVDPHYPAVCAHWGPRPLRTRAGRLVCVAITLVLL 160
DB 14 GVCAHRLH-----RGVYLMAVSVDPHYPAVCAHWGPRPLRTRAGRLVCVAITLVLL 67

OY 161 QTPLLLPMTKPLVGLKACMEYSSVESVGLPLMVLVAFATGFCGPVGIILSCYMKITW 220
DB 68 QTPLLLPMTKPLVGLKACMEYSSVESVGLPLMVLVAFATGFCGPVGIILSCYMKITW 110
OY 221 KLCST-----AREDPVTSRKGRHWRG 241
DB 111 LLWASGHPVLLVEDHLEAVQHSWENPVTSGKHRRG 148
RESULT 10
AAAY90652
ID AAAY90652 standard; Protein; 339 AA.
XX
XX AAAY90652;
XX 21-AUG-2000 (first entry)
XX Human mutant G protein-coupled receptor GPR17 (V234K).
XX G protein-coupled receptor; GPCR; constitutively active;
KW intracellular loop 3; transmembrane domain 6; drug screening;
KW agonist; antagonist; mutant; mutein.
XX
XX Homo sapiens.
XX Synthetic.
XX WO200022129-A1.
XX 20-APR-2000.
XX 12-OCT-1999; 99WO-US23938.
XX 13-OCT-1998; 98US-0170496.
XX (AREN-) ARENA PHARM INC.
XX Behan DP, Chalmers DT, Liaw CW;
PI
XX WPI; 2000-329165/28.
XX N-PSDB; AAA30718.
XX
XX Non-endogenous constitutively activated human G protein-coupled
PT receptors, useful for identifying agonists for use as pharmaceutical
PT agents -
XX
XX Example 2; Page 233-234; 341pp; English.
XX
XX The invention relates to constitutively active, non-endogenous versions
CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAAY90643-
CC AAAY90677 and AAAY90693-Y90687), and to DNA encoding them (AAA30709-A30743
CC and AAA30775-A30779). The mutant proteins of the invention contain a
CC mutation in a portion of the protein comprising intracellular loop 3
CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC is substituted for an endogenous residue in IC3 at a position 16 amino
CC acids N-terminal of an endogenous proline in TM6 to form a sequence
CC X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
CC or Ala, and is preferably Lys. When the endogenous residue at this
CC position is Lys, this residue is replaced by His, Arg or preferably Ala.
CC The 15 amino acid stretch between the substituted amino acid and the Pro
CC may be endogenous, non-endogenous, or a mixture of endogenous and
CC non-endogenous residues. The constitutively active GPCRs are useful for
CC identifying antagonists, agonists and partial agonists for use as
CC pharmaceutical agents. The mutant proteins are also useful in research
CC settings for elucidating the roles of the receptors in normal and
CC diseased conditions. Antagonists for a particular GPCR are useful for
CC treating diseases and disorders associated with that receptor. Because
CC the novel mutant GPCRs are constitutively active, they can be used
CC directly for screening of compounds without the need for endogenous
CC ligands. Sequences AAAY90643- AAAY90677 and AAAY90683-Y90687 the mutant
CC human GPCRs of the invention.
XX
XX Sequence 339 AA;
SQ

Query Match 21.1%; Score 395.5; DB 21; Length 339;
Best Local Similarity 30.5%; Pred. No. 2e-33;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;

QY 36 FYTALLVFSALGNLALCLTCOKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCVGLSS 95
DB 37 FYLLDFILALVGNLTALWLFIRDHKSGTPANVFLMHLAVADLSCLVLETRLVYHPSGNH 96

QY 96 WPFKGKLCRLTAFLVLYTDTYGGVYLMACVSDHYPAVCAHMGPRLTRAGARLVCAVAIW 155
DB 97 WPFGEIACRLTGLFLYLNMYASYIFLTCTISADRFIAIHPVKSLKRRPLVLAHLACAFW 156

QY 156 TLVLQTMPLLLMPKPLVGLKACME-YSSMESVLGLPLMWLVAIPAIGFCGPGVGIILSC 214
DB 157 VVAVAMAPLLVSPQTQTNHTVCLQYREKAS-----HVALVSLAVAFTPFTTTC 211

QY 215 YMKITWKLCTAREDPVTSRGRHWRGCLLTLLMLVAVVVCFSYHNLNIQFVARGMLHL 274
DB 212 YLLIIRSLRQGLRVEKRLTKAKR-----MIAVLAIFLVCFVPHVNRSVVLYHRSHG 266

QY 275 PSCAERRAFLLSLOATVALMMNCGTPIIYFFASTHYRKLWLLGIL---KLKG---SSSS 328
DB 267 ASCATORILANRITSCUTSLNGALDPIMYFFVAKFRHALCNLLCGKRLKGPFPFEG 326

QY 329 SSSSSSTPGKA 339
DB 327 KTNESLSAKS 337

RESULT 11
AAR53752
ID AAR53752 standard; Protein; 339 AA.
AC AAR53752;
XX
XX 25-MAR-2003 (updated)
DT 07-FEB-1995 (first entry)
XX
DE Seven transmembrane receptor (R12).
KW Primer; seven transmembrane receptor; receptor; amplification.
XX
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 33..57
FT Domain /label= Transmembrane domain.
FT Domain 68..90
FT Domain /label= Transmembrane domain.
FT Domain 106..127
FT Domain /label= Transmembrane domain.
FT Domain 145..168
FT Domain /label= Transmembrane domain.
FT Domain 193..217
FT Domain /label= Transmembrane domain.
FT Domain 233..251
FT Domain /label= Transmembrane domain.
FT Domain 290..312
FT Domain /label= Transmembrane domain.
XX
XX WO9412635-A2.
XX
XX 09-JUN-1994.
XX
XX 17-NOV-1993; 93WO-US11153.
XX
XX 17-NOV-1992; 92US-0977452.
XX
XX (ICOS-) ICOS CORP.
XX
XX Godiska R, Gray PW, Schweickart VL;
XX
XX WPI; 1994-200264/24.

DR N-PSDB; AAQ66178.
XX
XX DNA encoding seven transmembrane receptors - used to develop
PT prods. for use as therapeutic or diagnostic agents for conditions
PT involving the receptors.
XX
XX Example 10; Page 81-82; 100pp; English.
XX

CC PCR using two primers (AAQ66174, AAQ66175) was performed to amplify a
CC partial sequence of the R20 seven transmembrane coding sequence
CC which was later used as a probe for isolating the R20 genomic clone
CC (AAQ66176) from a human placenta DNA library. During the isolation
CC of the R20 gene, two weakly hybridising sequences were identified
CC which had significant homology to other seven transmembrane
CC receptors. The probe was used to screen a human genomic foetal
CC liver DNA library, and while the R20 gene could not be identified in
CC this library, several weakly hybridising clones were plaque
CC purified, subcloned and sequenced. The two clones were designated
CC R2 (AAQ66177) and R12. This is the sequence encoded by the R12 clone.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 339 AA;
SQ

Query Match 20.9%; Score 391.5; DB 15; Length 339;
Best Local Similarity 30.5%; Pred. No. 5.4e-33;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;

QY 36 FYTALLVFSALGNLALCLTCOKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCVGLSS 95
DB 37 FYLLDFILALVGNLTALWLFIRDHKSGTPANVFLMHLAVADLSCLVLETRLVYHPSGNH 96

QY 96 WPFKGKLCRLTAFLVLYTDTYGGVYLMACVSDHYPAVCAHMGPRLTRAGARLVCAVAIW 155
DB 97 WPFGEIACRLTGLFLYLNMYASYIFLTCTISADRFIAIHPVKSLKRRPLVLAHLACAFW 156

QY 156 TLVLQTMPLLLMPKPLVGLKACME-YSSMESVLGLPLMWLVAIPAIGFCGPGVGIILSC 214
DB 157 VVAVAMAPLLVSPQTQTNHTVCLQYREKAS-----HVALVSLAVAFTPFTTTC 211

QY 215 YMKITWKLCTAREDPVTSRGRHWRGCLLTLLMLVAVVVCFSYHNLNIQFVARGMLHL 274
DB 212 YLLIIRSLRQGLRVEKRLTKAKR-----MIAVLAIFLVCFVPHVNRSVVLYHRSHG 266

QY 275 PSCAERRAFLLSLOATVALMMNCGTPIIYFFASTHYRKLWLLGIL---KLKG---SSSS 328
DB 267 ASCATORILANRITSCUTSLNGALDPIMYFFVAKFRHALCNLLCGKRLKGPFPFEG 326

QY 329 SSSSSSTPGKA 339
DB 327 KTNESLSAKS 337

RESULT 12
AAW07617
ID AAW07617 standard; Protein; 339 AA.
XX
XX AAW07617;
XX
XX 28-FEB-1997 (first entry)
DT
DE Human G-protein thrombin-like receptor.
XX
XX
KW G-protein; thrombin; receptor; diagnosis; therapy; agonist;
KW antagonist; haemophilia; wound healing; restenosis; angina;
KW inflammation.
XX
XX Homo sapiens.
XX
XX WO9639438-A1.
XX
XX 12-DEC-1996.
XX
XX 06-JUN-1995; 95WO-US07180.
PF

Db 267 ASCATORILANRITSLTSLNGALDPIWFFVAEKFRHALCNLCCKRLKGGPPSPFEG 326
QY 329 SSSSSSTPGKA 339
Db 327 KINESLSAKS 337

RESULT 14
AAB21697
ID AAB21697 standard; Protein; 339 AA.
XX AAB21697;
AC AAB21697;
DT 26-JAN-2001 (first entry)
XX Human 7TM receptor R12 cDNA clone protein.
DE Seven transmembrane receptor; 7TM; heptahelical; serpentine;
KW G-protein-coupled; V28; V31; V112; R20; R2; R12; RM3; gene therapy;
KW cancer.
XX Homo sapiens.
XX US6107475-A.
XX 22-AUG-2000.
XX 26-APR-1999; 99US-0299843.
XX 17-MAY-1994; 94US-0245242.
PR 01-JUN-1998; 98US-0088337.
PR 17-NOV-1992; 92US-0977452.
PR 17-NOV-1993; 93US-0153848.
XX (ICOS-) ICOS CORP.
XX Schwickart VL, Gray PW, Godiska R;
XX WPI; 2000-571335/53.
DR N-PSDB; AAA91725.
XX Polynucleotide encoding seven transmembrane receptors, antibody
PT specific to the receptor, agonist and antagonist of the receptor useful
PT for treating inflammation in a mammal -
XX Example 10; Columns 93-96; 61pp; English.

XX The present sequence is a novel seven transmembrane (7TM) receptors
CC (also known as heptahelical, serpentine or G-protein-coupled receptors).
CC The coding sequence for the present sequence may be used for gene
CC therapy for diseases such as cancer.

Query Match 20.9%; Score 391.5; DB 21; Length 339;
Best Local Similarity 30.5%; Pred.No. 5.4e-33;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;

QY 36 FYTALLVFSALGNLIALCTCQSKINCTGIVLHVAVSDLLFTVALPGRVVCYVLGSS 95
Db 37 FYLLDFILALVGNLALWLFIRDHKSGETPANVFLMLHVAADLSCLVLPRLVHFSGNH 96
QY 96 WPFKGKGLRLTAFLVYTDYGGVYLMACVSDHYPAVVCAGWGPRLTAGRRLVCVAIW 155
Db 97 WPFGEIACRLTGFLLYLMNYASIIYFLTCISADRLAIVHPVKSLLRRPYLHACAPLW 156
QY 156 TLVLLQTMPLLMPTKPLVGLKACME-YSSMESVLGLPLMLVAFAGFCGPGVGLISC 214
Db 157 VVAVAMAPLLVSPQTQNTHTVCLQLYREKAS-----HHALVSLAVAFTFPFTTVC 211
QY 215 YNKITWKLCTAREDPVTSRGRHWRCGLLTLLMLVAVVVCPSYHLNIKQFMAGMLHL 274
Db 212 YLLIIRSLRQGLRVE-----KRLKTKAVRMIATVLAFLVCFVPHVNRSVYLVHLHRSRG 256

QY 275 PSCAERRAPLLSQATVALMNNCGITPIIYFFASTHYRKMLLGIL---KLKG---SSSS 328
Db 267 ASCATORILANRITSLTSLNGALDPIWFFVAEKFRHALCNLCCKRLKGGPPSPFEG 326
QY 329 SSSSSSTPGKA 339
Db 327 KINESLSAKS 337

RESULT 15
AAY90618
ID AAY90618 standard; Protein; 339 AA.
XX AAY90618;
AC AAY90618;
DT 21-AUG-2000 (first entry)
XX Human G protein-coupled receptor GPR17.
DE G protein-coupled receptor; GPCR; constitutively active;
KW intracellular loop 3; transmembrane domain 6; drug screening;
KW agonist; antagonist.
XX Homo sapiens.
XX WO2000022129-A1.
XX 20-APR-2000.
XX 12-OCT-1999; 99WO-US23938.
XX 13-OCT-1998; 98US-0170496.
XX (AREN-) ARENA PHARM INC.
XX Behan DP, Chalmers DT, Liaw CM;
XX WPI; 2000-329165/28.
DR N-PSDB; AAA30601.
XX Non-endogenous constitutively activated human G protein-coupled
PT receptors, useful for identifying agonists for use as pharmaceutical
PT agents -
XX Example 1; Page 126-127; 341pp; English.

XX The invention relates to constitutively active, non-endogenous versions
CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
CC and AAA30775-A30779). The mutant proteins of the invention contain a
CC mutation in a portion of the protein comprising intracellular loop 3
CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC is substituted for an endogenous residue in IC3 at a position 16 amino
CC acids N-terminal of an endogenous proline in TM6 to form a sequence
CC X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
CC or Ala, and is preferably Lys. When the endogenous residue at this
CC position is Lys, this residue is replaced by His, Arg or preferably Ala.
CC The 15 amino acid stretch between the substituted amino acid and the Pro
CC may be endogenous, non-endogenous, or a mixture of endogenous and
CC non-endogenous residues. The constitutively active GPCRs are useful for
CC identifying antagonists, agonists and partial agonists for use as
CC pharmaceutical agents. The mutant proteins are also useful in research
CC settings for elucidating the roles of the receptors in normal and
CC diseased conditions. Antagonists for a particular GPCR are useful for
CC treating diseases and disorders associated with that receptor. Because
CC the novel mutant GPCRs are constitutively active, they can be used
CC directly for screening of compounds without the need for endogenous
CC ligands. The present sequence represents a human wild-type GPCR referred
CC to in an exemplification of the invention.

XX Sequence 339 AA;
SQ

```
Query Match      20.9%; Score 391.5; DB 21; Length 339;
Best Local Similarity 30.5%; Pred No. 5.4e-33;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;

QY 36 FYTALLVFSALGNILALCLTCKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCVVLGSS 95
Db 37 FYLLDPILALVGNLTALWLFIRDHKSQTPANVFLMHLAVADLSCVLVLPTRLVVHFSGNH 96
QY 96 WPFKGKLCRLTAFVLYTDTYGGVYLMACVSDHYPAVCAHWGPRLTAGRARLVCVALIW 155
Db 97 WPFGEIACRLTGFLFYLNMYAIYFTCTISADRFIAIVHPVKSILRRPLAHLACAFIW 156
QY 156 TLVLLQTMPELLMPTKPLVGLKACME-YSSMESVLGLPLMVLVAFICGCPVGIILSC 214
Db 157 VVVAVAMAPLLVSPQTVQTNHTVCLQLYREKAS-----HHALVSLAVAFPTPFTTITVC 211
QY 215 YMKITWKLCTAREDPVTSRKGHRWGCLLLMLLVAVVVCPSPYHLAIKOFMARGMLHL 274
Db 212 YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIVLAIFLVCFVPHVNRSVYVLHYRSHG 266
QY 275 PSCAERRAFLLSQATVALMNNVNCGITPIIYFPFASHTYRKWLLGIL---KIMG---SSSS 328
Db 267 ASCATQRIILANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGGPPPPFEG 326
QY 329 SSSSSSTPCKA 339
Db 327 KTNESLSAKS 337
```

Search completed: February 19, 2004, 19:33:00
Job time : 45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 19:32:10 ; Search time 21 Seconds
(without alignments)
725.329 Million cell updates/sec

Title: US-10-041-615-34
Perfect score: 1876
Sequence: 1 MIKLGPPYPTTKIKTKID.....SETPSITQARGSMFLAEHV 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata/1/aa/5A_COMB.pep.*
2: /cgn2_6/prodata/1/aa/5B_COMB.pep.*
3: /cgn2_6/prodata/1/aa/6A_COMB.pep.*
4: /cgn2_6/prodata/1/aa/6B_COMB.pep.*
5: /cgn2_6/prodata/1/aa/PCITUS_COMB.pep.*
6: /cgn2_6/prodata/1/aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	635.5	33.9	348	US-08-852-824-17	Sequence 17, Appli
2	632.5	33.7	361	US-08-383-750-4	Sequence 4, Appli
3	632.5	33.7	361	US-08-352-678-4	Sequence 4, Appli
4	632.5	33.7	361	US-09-536-954-4	Sequence 4, Appli
5	632.5	33.7	361	US-09-170-496D-78	Sequence 78, Appli
6	632.5	33.7	361	PCT-US93-09635-4	Sequence 4, Appli
7	630.5	33.6	361	US-09-170-496B-206	Sequence 206, Appli
8	395.5	21.1	339	US-09-170-496D-182	Sequence 182, Appli
9	391.5	20.9	339	US-08-153-848-44	Sequence 44, Appli
10	391.5	20.9	339	US-08-812-871-3	Sequence 3, Appli
11	391.5	20.9	339	US-09-299-843A-44	Sequence 44, Appli
12	391.5	20.9	339	US-09-088-337B-44	Sequence 44, Appli
13	391.5	20.9	339	US-09-170-496B-32	Sequence 32, Appli
14	391.5	20.9	339	PCT-US93-11153-44	Sequence 44, Appli
15	391.5	20.9	339	PCT-US95-07180-2	Sequence 2, Appli
16	353	18.8	357	PCT-US95-07180-2	Sequence 3, Appli
17	340	18.1	425	US-08-313-553-13	Sequence 13, Appli
18	340	18.1	425	US-08-767-993-13	Sequence 13, Appli
19	339.5	18.1	425	US-07-657-769B-69	Sequence 69, Appli
20	339.5	18.1	425	US-08-097-938-7	Sequence 7, Appli
21	339.5	18.1	425	US-07-789-184-220	Sequence 220, Appli
22	339.5	18.1	425	US-08-476-000-7	Sequence 7, Appli
23	339.5	18.1	425	US-08-475-263-220	Sequence 220, Appli
24	339.5	18.1	425	US-08-472-840-7	Sequence 7, Appli
25	339.5	18.1	425	US-08-485-886-220	Sequence 220, Appli
26	339.5	18.1	425	US-08-477-362-220	Sequence 220, Appli
27	339.5	18.1	425	US-08-477-134-220	Sequence 220, Appli

RESULT 1
US-08-852-824-17
; Sequence 17, Application US/08852824C
; Patent No. 6060272
; GENERAL INFORMATION:
; APPLICANT: Li et al.
; TITLE OF INVENTION: Human G-Protein Coupled Receptors
; FILE REFERENCE: 1488.1220000
; CURRENT APPLICATION NUMBER: US/08/852,824C
; CURRENT FILING DATE: 1997-05-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 17
; LENGTH: 348
; TYPE: PRT
; ORGANISM: genomic
US-08-852-824-17

ALIGNMENTS

Query Match 33.9%; Score 635.5; DB 3; Length 348;
Best Local Similarity 38.5%; Pred. No. 3.6e-39;
Matches 132; Conservative 69; Mismatches 139; Indels 3; Gaps 2;

Qy	2	IKLGPPYPTTKIKTKIDLNVESQTKIFLRSIFLTALIVFSALGNILALCLTCQSRK	61	Sequence 3, Appli
Db	1	IQMANPTPSATPONDCCDIYAHSTARIYK-PLHYSLVFIIGLVGNLLALVIVQNRKK	59	Sequence 7, Appli
Qy	62	INCTGIYLVHLAVSDLLFTVALPGRVVYVYLGSSWPGKGLCRITAFVLTDTYGGVYLM	121	Sequence 57, Appli
Db	60	INSTLYSTNLVISDILFTTALPTRIAYANGDFMRIGDLCRITALVFYINTYAGVNF	119	Sequence 220, Appli
Qy	122	ACTVSDHYPAVCAHWGPRRTAGRAELVCVAIWTLLQTMPLLLMPTKPLVGLKACM	181	Sequence 7, Appli
Db	120	TCLSIDRFIAVHPLRYNKRIEHAQVCIFWILVFAQTFLPINPMSKQEAERTCM	179	Sequence 220, Appli
Qy	182	EYSWMSVYGLPLMLVAFIGCPGVGIIILSCYMKITWKLCSSTAREDPVTSRGRHWRG	241	Sequence 3, Appli
Db	180	EYPNFEETKSPWILLGACPIGVYPLIIIIKICYSQICCKLFRITAKONPLTEKSGVYKKA	239	Sequence 7, Appli
Qy	242	CLLTLLMLVAVVCFSPHYLNKQFMARGM--LHLPSCARRAFLLSLQATVALMMNCG	299	Sequence 220, Appli
Db	240	LNTILIIIVFVLCFTPYHVAIIQHMIKLRFNFLECSQRHSFQISLHTVCLMNFCC	299	Sequence 220, Appli
Qy	300	ITPIIYFFASTHYRKWLLGLIKLKGSSSSSSSSSSSTFGKASSE	342	Sequence 56, Appli
Db	300	MDPIIYFPACKGYKRYMRLKQVSVSISSAVKSAPEENSRE	342	Sequence 56, Appli

RESULT 2
US-08-383-750-4
; Sequence 4, Application US/08383750

Patent No. 5744301
 GENERAL INFORMATION:
 APPLICANT: Birkenbach, Mark
 APPLICANT: Kieff, Elliott
 TITLE OF INVENTION: Epstein Barr Virus Induced Genes
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 STREET: 1100 New York Avenue, N.W.,
 STREET: Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20005-3934

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/383,750
 FILING DATE: Herewith

CLASSIFICATION: 436
 ATTORNEY/AGENT INFORMATION:
 NAME: Fox, Samuel, L.
 REGISTRATION NUMBER: 30,353
 REFERENCE/DOCKET NUMBER: 0627.3300001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 361 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-383-750-4

Query Match 33.7%; Score 632.5; DB 1; Length 361;
 Best Local Similarity 37.8%; Pred. No. 6.2e-39;
 Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;
 QY 2 IKLGPFYTPPKI-KTKIKDLNVESTQVKIPLSLFTALLVFSALGNILALCLTCQKSR 60
 DB 3 IOMANNFTPSATPQNGDCDLYAHSTARIWM-PLHYSLVFIIGLVGNLLAVVIVQNRK 61
 QY 61 KINCTGIYVHLAVSLLFTVALPGRVVCVVLGSSWPFKGLCRLTAFVLYTDTYGGVYL 120
 DB 62 KINSTLYSTNLVSDILFTTALPTRIAYYANGFDWIGDALCRITALTALVFIINTYAGVNF 121
 QY 121 MACVSDHYPAVCAHNGPRLRTAGRLVCVAITWLVLTQTPMLLMPMTKPLVGLKAC 180
 DB 122 MTCLSIDRFIAVHPLRYNKIKRIEHAHKGVCIFWILVFAQTLPPLINPMSQBAERITC 181
 QY 181 MEYSMESVGLPLMLVAFAGFCGPGVGIILSCYMKITWKLCTAREDPVTSRKGHRWR 240
 DB 182 MEYPNFEETKSLPWLIGACFVGLVPLIILICYSQICCKLFRITAKQNPLETEKSGVNNK 241
 QY 241 GCLLTLLMLVAVVCFSPYHNLINQFARMG--LHLPSCAERRAFILSIQATVALMMNMC 298
 DB 242 ALNTILILIVVFLCFTPYHVAIIQHKIKRPFNFLECSQSHSFQISLHFTVCLMNFNC 301
 QY 299 GITPIIYFFASTHYRKLGLILKLGSSSSSSSSSSSTPGKAS---SETPSITQAGS 352
 DB 302 CMDPPIYFFACKYKRWMLKQVSVSISSAVKSAPEENSREMTETQWMIHKS 358

RESULT 3
 US-08-352-678-4
 Sequence 4, Application US/08352678
 Patent No. 6043351
 GENERAL INFORMATION:
 APPLICANT: Birkenbach, Mark
 APPLICANT: Kieff, Elliott

TITLE OF INVENTION: EPSTEIN BARR VIRUS INDUCED GENES
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
 STREET: 600 Atlantic Avenue
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02210
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/352,678
 FILING DATE: 30-NOV-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/980,518
 FILING DATE: 25-NOV-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Gates, Edward R.
 REGISTRATION NUMBER: 31,616
 REFERENCE/DOCKET NUMBER: E0801/7044
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-720-3500
 TELEFAX: 617-720-2441
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 361 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-352-678-4

Query Match 33.7%; Score 632.5; DB 3; Length 361;
 Best Local Similarity 37.8%; Pred. No. 6.2e-39;
 Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;
 QY 2 IKLGPFYTPPKI-KTKIKDLNVESTQVKIPLSLFTALLVFSALGNILALCLTCQKSR 60
 DB 3 IOMANNFTPSATPQNGDCDLYAHSTARIWM-PLHYSLVFIIGLVGNLLAVVIVQNRK 61
 QY 61 KINCTGIYVHLAVSLLFTVALPGRVVCVVLGSSWPFKGLCRLTAFVLYTDTYGGVYL 120
 DB 62 KINSTLYSTNLVSDILFTTALPTRIAYYANGFDWIGDALCRITALTALVFIINTYAGVNF 121
 QY 121 MACVSDHYPAVCAHNGPRLRTAGRLVCVAITWLVLTQTPMLLMPMTKPLVGLKAC 180
 DB 122 MTCLSIDRFIAVHPLRYNKIKRIEHAHKGVCIFWILVFAQTLPPLINPMSQBAERITC 181
 QY 181 MEYSMESVGLPLMLVAFAGFCGPGVGIILSCYMKITWKLCTAREDPVTSRKGHRWR 240
 DB 182 MEYPNFEETKSLPWLIGACFVGLVPLIILICYSQICCKLFRITAKQNPLETEKSGVNNK 241
 QY 241 GCLLTLLMLVAVVCFSPYHNLINQFARMG--LHLPSCAERRAFILSIQATVALMMNMC 298
 DB 242 ALNTILILIVVFLCFTPYHVAIIQHKIKRPFNFLECSQSHSFQISLHFTVCLMNFNC 301
 QY 299 GITPIIYFFASTHYRKLGLILKLGSSSSSSSSSSSTPGKAS---SETPSITQAGS 352
 DB 302 CMDPPIYFFACKYKRWMLKQVSVSISSAVKSAPEENSREMTETQWMIHKS 358

RESULT 4
 US-09-536-954-4
 Sequence 4, Application US/09536954
 Patent No. 6500926
 GENERAL INFORMATION:
 APPLICANT: Birkenbach, Mark
 APPLICANT: Kieff, Elliott
 TITLE OF INVENTION: EPSTEIN BARR VIRUS INDUCED GENES

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,954
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/352,678
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gates, Edward R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: B0801/7044
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-536-954-4

Query Match 33.7%; Score 632.5; DB 4; Length 361;
Best Local Similarity 37.8%; Pred. No. 6.2e-39;
Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;

QY 2 IKLGPVFTPTKI-KTKIKDLNVESTQVTKIFLRSFYTALLVFSALGNILALCLTCQKSR 60
DB 3 IQANNFTPTSATPQGNDCDLVAHSTARIWM-PLHYSLVFIIGLVGNLLALVWVQNRK 61
QY 61 KINCTGIYLVHLSVLDLFTVALPGRVVCYVIGSSWPFQKGLCRITAFVLYTDYGGVYL 120
DB 62 KINSTLTSTNLVSDILFTTALPTRIAYANGFDWRIGDALCRITAFVLYTAYGVNF 121
QY 121 MACVSDHYPAVCAHWGRLTAGRLVCVAIWTLVLTQMPLLMPMTKPLVVKLAC 180
DB 122 MTCLSDIRFIADVHPRLYNKRIEHAQVCIFWILVPAQILPLINPMSQAEERITC 181
QY 181 MEYSSMESVGLPLMLVFAIGFCGPVGIILSCYNMKITWKLCTAREDPVTSRGRHWR 240
DB 182 MEYPNFEETKSLPWILLGACFIGYVLPILIIICYSQICKLFRITAKONPLTEKSGVNNK 241
QY 241 GCLLTLLMLVAVVCPSPHNLKQPMARMG--LHLPSCAERRAFLLSLQATVALMMNMC 298
DB 242 ALNTIILIIIVFVLCFTPHVAIQHMKIKLAFSNFLECSQRHSFOISLHFTVCLNFC 301
QY 299 GITPIYFFASTHYRKMILGILKKGSSSSSSSSSTPGKAS---SETPSITQARGS 352
DB 302 CMDPFIYFFACKYKRWMLKQVSVSISSAVKSAPEENSREMTETQMWIHSKSS 358

RESULT 5
US-09-170-496D-78
Sequence 78, Application US/09170496D
Patent No. 655339
GENERAL INFORMATION:
APPLICANT: Behan, Dominic P.
APPLICANT: Chalmers, Derek T.
TITLE OF INVENTION: No. 655339-Endogenous, Constitutively Activated Human G Protein-

TITLE OF INVENTION: Receptors
FILE REFERENCE: AREN-0040
CURRENT APPLICATION NUMBER: US/09/170,496D
CURRENT FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 294
SOFTWARE: PatentIn version 3.1
SEQ ID NO 78
LENGTH: 361
TYPE: PRT
ORGANISM: Homo sapiens
US-09-170-496D-78

Query Match 33.7%; Score 632.5; DB 4; Length 361;
Best Local Similarity 37.8%; Pred. No. 6.2e-39;
Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;

QY 2 IKLGPVFTPTKI-KTKIKDLNVESTQVTKIFLRSFYTALLVFSALGNILALCLTCQKSR 60
DB 3 IQANNFTPTSATPQGNDCDLVAHSTARIWM-PLHYSLVFIIGLVGNLLALVWVQNRK 61
QY 61 KINCTGIYLVHLSVLDLFTVALPGRVVCYVIGSSWPFQKGLCRITAFVLYTDYGGVYL 120
DB 62 KINSTLTSTNLVSDILFTTALPTRIAYANGFDWRIGDALCRITAFVLYTAYGVNF 121
QY 121 MACVSDHYPAVCAHWGRLTAGRLVCVAIWTLVLTQMPLLMPMTKPLVVKLAC 180
DB 122 MTCLSDIRFIADVHPRLYNKRIEHAQVCIFWILVPAQILPLINPMSQAEERITC 181
QY 181 MEYSSMESVGLPLMLVFAIGFCGPVGIILSCYNMKITWKLCTAREDPVTSRGRHWR 240
DB 182 MEYPNFEETKSLPWILLGACFIGYVLPILIIICYSQICKLFRITAKONPLTEKSGVNNK 241
QY 241 GCLLTLLMLVAVVCPSPHNLKQPMARMG--LHLPSCAERRAFLLSLQATVALMMNMC 298
DB 242 ALNTIILIIIVFVLCFTPHVAIQHMKIKLAFSNFLECSQRHSFOISLHFTVCLNFC 301
QY 299 GITPIYFFASTHYRKMILGILKKGSSSSSSSSSTPGKAS---SETPSITQARGS 352
DB 302 CMDPFIYFFACKYKRWMLKQVSVSISSAVKSAPEENSREMTETQMWIHSKSS 358

RESULT 6
PCT-US93-09636-4
Sequence 4, Application PC/TUS9309636
GENERAL INFORMATION:
APPLICANT: Birkenbach, Mark
TITLE OF INVENTION: Epstein Barr Virus Induced Genes
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/09636
FILING DATE: herewith
CLASSIFICATION:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 amino acids
TYPE: amino acid
TOPOLOGY: linear

Qy	2	IKLGPFYTPPTKI-KTKIKDINVESQTVKIFLSLFTYATLLVPSALGNIALCITCQKSR	60
Db	3	IQMANFTPPSATPGQNDCLDYAHSHSTARIW-PLHSLVFIIGLVGNLLALVTVQNRK	61
Qy	61	KINCTGIYLVHLVSDLLFTVALPGRVVCVGLSSNPFPGKGCLETAFLVLTDTTGGVYL	120
Db	62	KINSTTLTSLTSLVSDILFTTALPRTAYYAMPGRIGDLCRIITAFVYINTVAGVNF	121
Qy	121	MACVSVDHYPAVVCVCAHMGFRLTAGRARLCVCAIWTLLVLTQTMPLLMPTMKFLGKIAC	180
Db	122	MTCLSIDRFIAVHPRLYRKIRIEAHKGVCIFWVLVPAQTLFLLINPMSKQEAERTC	181
Qy	181	MEYSSWESVGLPLMVLVAFAGCGPVGIIILCSYWKITWKLCSATAREDPVTSRGRHWR	240
Db	182	MEYPNFEETKSLFWLLIGACFIVGLVPLIIILICYSQICCKLFRTAKOMPLTEKSGVNKK	241
Qy	241	GCLLFTLLMLVAWVCFSPLHNLINIKQFMARG--LHLPSCAERRAFLLSLQATVALMMNVC	298

RESULT 9
US-08-153-848-44
; Sequence 44, Application US/08153848
; Patent No. 5759804
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweickart, Vicki L.
; TITLE OF INVENTION: Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago

STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/153,848
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 5759804and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 339 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-153-848-44

Query Match 20.9%; Score 391.5; DB 1; Length 339;

Best Local Similarity 30.5%; Pred. No. 2.1e-21;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;

Qy	36	FTYALLVFSALGNIALCLTCQSKINCTGYLVHVLAVSDLLFTVALPGRVVCVVLGSS	95
Db	37	FYLLDFILAVGNLTALWLFIRDHKSGTPANVFLMHLAVADISCVLVLPRLVHFSGNH	96
Qy	96	WPGKGLCRITAFVLYTDTYGGVYLACVSDHYPAVCAHWGPRRTAGRARLVCAVIM	155
Db	97	WPFGEIACRLTGFLFYLNMVASYIFLTCSADRLAIVHPVKSLKRRPLYAHLACAPLM	156
Qy	156	TLVLLQTMPLLLMPMTKPLVGLKACME-YSSMESVLGLPLMLVAPAFGCGPVGIIISC	214
Db	157	VVAVAMAPLLVSPQVTQNTHTVVCQLYREKAS-----HVALVSLAVAFTPFITVTTC	211
Qy	215	YMKITWKLCTAREDPVTSRGRHWRGCLLTLLMLVAVVCFSPYHLNLIKQFMARGMLHL	274
Db	212	YLLIIRSLRQGLRVE-----KRLTKAVRMIAIVLAIFLVCFVYHVNRSVYVLYHRSHG	266
Qy	275	PSCAERRAFLLSLQATVALMNMCGITPIIYFFASTHYRKWLLGIL---KLKG---SSSS	328
Db	267	ASCATORILANRITSLTSLNGALDPIWFFVFAEKFRHALCNLLCGKRLKGPSPFEG	326
Qy	329	SSSSSSTPGKA	339
Db	327	KTNESLSAKS	337

RESULT 10

US-08-812-871-3
Sequence 3, Application US/08812871
Patent No. 595303
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Guedler, Karl
APPLICANT: Muzong Cheng
TITLE OF INVENTION: NOVEL HUMAN CHEMOKINE RECEPTOR-LIKE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/812,871
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0237 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 339 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 992700
US-08-812-871-3

Query Match 20.9%; Score 391.5; DB 2; Length 339;

Best Local Similarity 30.5%; Pred. No. 2.1e-21;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;

Qy	36	FTYALLVFSALGNIALCLTCQSKINCTGYLVHVLAVSDLLFTVALPGRVVCVVLGSS	95
Db	37	FYLLDFILAVGNLTALWLFIRDHKSGTPANVFLMHLAVADISCVLVLPRLVHFSGNH	96
Qy	96	WPGKGLCRITAFVLYTDTYGGVYLACVSDHYPAVCAHWGPRRTAGRARLVCAVIM	155
Db	97	WPFGEIACRLTGFLFYLNMVASYIFLTCSADRLAIVHPVKSLKRRPLYAHLACAPLM	156
Qy	156	TLVLLQTMPLLLMPMTKPLVGLKACME-YSSMESVLGLPLMLVAPAFGCGPVGIIISC	214
Db	157	VVAVAMAPLLVSPQVTQNTHTVVCQLYREKAS-----HVALVSLAVAFTPFITVTTC	211
Qy	215	YMKITWKLCTAREDPVTSRGRHWRGCLLTLLMLVAVVCFSPYHLNLIKQFMARGMLHL	274
Db	212	YLLIIRSLRQGLRVE-----KRLTKAVRMIAIVLAIFLVCFVYHVNRSVYVLYHRSHG	266
Qy	275	PSCAERRAFLLSLQATVALMNMCGITPIIYFFASTHYRKWLLGIL---KLKG---SSSS	328
Db	267	ASCATORILANRITSLTSLNGALDPIWFFVFAEKFRHALCNLLCGKRLKGPSPFEG	326
Qy	329	SSSSSSTPGKA	339
Db	327	KTNESLSAKS	337

RESULT 11

US-09-299-843A-44
Sequence 44, Application US/09299843A
Patent No. 6107475
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: No. 6107475el Seven Transmembrane Receptors

NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/299,843A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/088,337
FILING DATE: 01-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,848
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Jill E. Uhl
REGISTRATION NUMBER: 43,213
REFERENCE/DOCKET NUMBER: 27866/32059B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX:
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 339 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-299-843A-44

Query Match 20.9%; Score 391.5; DB 3; Length 339;
Best Local Similarity 30.5%; Pred. No. 2.1e-21;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;
QY 36 FYTALLVPSALGNLALCLTQCKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
DB 37 FYLLDFILAVGNLTALWLFIRDHKSCTPANVFLMHLAVADLSCLVLPTRLVYHFSGNH 96
QY 96 WPFKGKGLRLTAFLVLTDTYGGVYLMACVSDHYPAVCAHWGPRLRRTAGRLVCAI 155
DB 97 WPFGEIACRLTGFLYLNMYASITFLTCTISADRFIAIVHPVKSUKLRRLPYAHLACAF 156
QY 156 TLVLQTMPLLMPTKPLVGLKACME-YSSMESVGLPLMLVAVVCFSPYHLNIPQFARGMLHL 214
DB 157 VVAVAMAPLLVSPQTQNTNHTVCLQYREKAS-----HHALVSLAVAFPTFTTVC 211
QY 215 YMKITWKLCTAREDPVTSRGRHWGCLLTLLMLVAVVCFSPYHLNIPQFARGMLHL 274
DB 212 YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIHLAIFLVCFVPHVNRVSVVLYHRS 266
QY 275 PCSAERRAFLLSQATVALMNNCGITPIIYFFASTHYRKLGLGIL---KLKG---SSSS 328
DB 267 ASCAQRIILANRITSLCTSLNGALDPIWFFVAEKFRHALCNLCGRKLGKPPSPFEG 326
QY 329 SSSSSSTPGKA 339
DB 327 KTNESLSAKS 337

RESULT 12

US-09-088-337B-44
Sequence 44, Application US/09088337B
Patent No. 6348574
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
Gray, Patrick W.
Schweikart, Vicki L.
TITLE OF INVENTION: No. 6348574e1 Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/088,337B
FILING DATE: 01-JUN-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,848
FILING DATE: 17-NOV-1993
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 6348574e1, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 339 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-088-337B-44

Query Match 20.9%; Score 391.5; DB 4; Length 339;
Best Local Similarity 30.5%; Pred. No. 2.1e-21;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;
QY 36 FYTALLVPSALGNLALCLTQCKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
DB 37 FYLLDFILAVGNLTALWLFIRDHKSCTPANVFLMHLAVADLSCLVLPTRLVYHFSGNH 96
QY 96 WPFKGKGLRLTAFLVLTDTYGGVYLMACVSDHYPAVCAHWGPRLRRTAGRLVCAI 155
DB 97 WPFGEIACRLTGFLYLNMYASITFLTCTISADRFIAIVHPVKSUKLRRLPYAHLACAF 156
QY 156 TLVLQTMPLLMPTKPLVGLKACME-YSSMESVGLPLMLVAVVCFSPYHLNIPQFARGMLHL 214
DB 157 VVAVAMAPLLVSPQTQNTNHTVCLQYREKAS-----HHALVSLAVAFPTFTTVC 211
QY 215 YMKITWKLCTAREDPVTSRGRHWGCLLTLLMLVAVVCFSPYHLNIPQFARGMLHL 274
DB 212 YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIHLAIFLVCFVPHVNRVSVVLYHRS 266
QY 275 PCSAERRAFLLSQATVALMNNCGITPIIYFFASTHYRKLGLGIL---KLKG---SSSS 328
DB 267 ASCAQRIILANRITSLCTSLNGALDPIWFFVAEKFRHALCNLCGRKLGKPPSPFEG 326
QY 329 SSSSSSTPGKA 339

[illegible]

Search completed: February 19, 2004, 19:35:29
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 19:33:05 ; Search time 40 Seconds
(without alignments)
1884.441 Million cell updates/sec

Title: US-10-041-615-34

Perfect score: 1876

Sequence: 1 MIKLGYPPTPTKTKIKTKIKO.....SETPSITQKRGSMFLAHVV 360

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/FCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1876	100.0	360	12	US-10-041-615-34
2	884	47.1	323	12	Sequence 34, Appl
3	884	47.1	323	12	Sequence 1620, Ap
4	721	38.4	229	12	Sequence 1294, Ap
5	635.5	33.9	348	9	Sequence 109, Appl
6	632.5	33.7	361	12	Sequence 17, Appl
7	632.5	33.7	361	12	Sequence 110, Appl
8	632.5	33.7	361	15	US-10-041-615-110
9	632.5	33.7	361	15	US-10-222-024-2
10	630.5	33.6	361	15	US-10-251-385-78
11	630.5	33.6	361	15	US-10-225-567A-112
12	395.5	21.1	339	15	Sequence 206, Appl
13	391.5	20.9	339	15	Sequence 95, Appl
14	391.5	20.9	339	9	Sequence 182, Appl
15	391.5	20.9	339	10	Sequence 12, Appl
					Sequence 2, Appl
					Sequence 4, Appl
					Sequence 8, Appl
					Sequence 4, Appl
					Sequence 12, Appl
					Sequence 14, Appl
					Sequence 112, Appl
					Sequence 32, Appl
					Sequence 6, Appl
					Sequence 6, Appl
					Sequence 11, Appl
					Sequence 103, Appl
					Sequence 113, Appl
					Sequence 25, Appl
					Sequence 198, Appl
					Sequence 21, Appl
					Sequence 80, Appl
					Sequence 4, Appl
					Sequence 3, Appl
					Sequence 324, Appl
					Sequence 4, Appl
					Sequence 130, Appl
					Sequence 12, Appl
					Sequence 56, Appl
					Sequence 10, Appl
					Sequence 18, Appl
					Sequence 1288, Ap
					Sequence 223, Appl
					Sequence 479, Appl
					Sequence 21, Appl

Query Match 100.0%; Score 1876; DB 12; Length 360;
Best Local Similarity 100.0%; Pred. No. 3.3e-164;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIKLGYPPTPTKTKIKTKIKLNVESQTKIFLSLFTALLVFSALGNILALCLTCOKSR 60
Db 1 MIKLGYPPTPTKTKIKTKIKLNVESQTKIFLSLFTALLVFSALGNILALCLTCOKSR 60
Qy 61 KINCTGTYLVHLAVSDLLFTVALPGRVVCVVLGSSWPFKGLCLRLTAFLVLYTDTGGVYL 120

ALIGNMENTS

RESULT 1

US-10-041-615-34
Sequence 34, Application US/10041615
Publication No. US20040014038A1
GENERAL INFORMATION:
APPLICANT: Casman, Stacie J
APPLICANT: Edinger, Shlomit R
APPLICANT: Ellerman, Karen
APPLICANT: Smithson, Glenda
APPLICANT: Kekuda, Ramesh
APPLICANT: Padigaru, Muralidhara
TITLE OF INVENTION: No. US20040014038A1 GPCR-Like Proteins and Nucleic Acids Encodi
FILE REFERENCE: 21402-233-061
CURRENT APPLICATION NUMBER: US/10/041,615
PRIOR FILING DATE: 2003-01-29
PRIOR APPLICATION NUMBER: 60/259,552
PRIOR FILING DATE: 2001-01-03
PRIOR APPLICATION NUMBER: 60/260,544
PRIOR FILING DATE: 2001-01-09
PRIOR APPLICATION NUMBER: 60/277,405
PRIOR FILING DATE: 2001-03-20
NUMBER OF SEQ ID NOS: 174
SOFTWARE: Curaseq version 0.1
SEQ ID NO 34
LENGTH: 360
TYPE: PRT
ORGANISM: Homo sapiens
US-10-041-615-34

Db 61 KINCTGIYVHLAVSOLLTVALPGRVVCVILGSSWPFGLCRITAFVLYTDTYGGVYL 120
Qy 121 MACVSDHYPAVCAHMGPRLTAGARLVCAIWTLLVLTQMPLLMPMTKPLVGLKAC 180
Db 121 MACVSDHYPAVCAHMGPRLTAGARLVCAIWTLLVLTQMPLLMPMTKPLVGLKAC 180
Qy 181 MEYSMESVGLPMLVAFAGFCGPGVGIILSCYNKTIWKLCSTAREDPVTSRGRHWR 240
Db 181 MEYSMESVGLPMLVAFAGFCGPGVGIILSCYNKTIWKLCSTAREDPVTSRGRHWR 240
Qy 241 GCLLTLLMLVAVVCFSPYHLNKKQWARGMLHLPSCAERRAFLLSLQATVALMMNCGI 300
Db 241 GCLLTLLMLVAVVCFSPYHLNKKQWARGMLHLPSCAERRAFLLSLQATVALMMNCGI 300
Qy 301 TPIIYFFASTHYRKWLGLILKLGSSSSSSSSSSSTPGKASSETPSITQARGSMFLAEHV 360
Db 301 TPIIYFFASTHYRKWLGLILKLGSSSSSSSSSSSTPGKASSETPSITQARGSMFLAEHV 360

RESULT 2
US-10-017-161-1620
; Sequence 1620, Application US/10017161
; Publication No. US2003014368A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; PRIOR FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1620
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-1620

Query Match 47.1%; Score 884; DB 12; Length 323;
Best Local Similarity 65.7%; Pred. No. 4.7e-73;
Matches 195; Conservative 8; Mismatches 36; Indels 58; Gaps 5;
Qy 101 GLCRLTAFVLYTDTYGGVILMACVSDHYPAVCAHMGPRLTAGARLVCAIWTLLV 160
Db 48 GVCALHRLH-----RGVYLMACVSDHYPAVCAHMGPRLTAGARLVCAIWTLLV 101
Qy 161 QTMPELLMPMTKPLVGLKACMEYSMESVGLGLPLMLVAFAGFCGPGVGIILSCYNKTIW 220
Db 102 QTMPELLMPMTKPLVGLKACMEYSMESVGLGLPLMLVAFAGFCGPGVGIILSCYNKTIW 161
Qy 221 KLCs-----TAREDPVTSRGRHWRGCLLTLLMLVAVVCFSPYHLNKKQFM----- 267
Db 162 KLCSTAGRTQAREDPVTSRGRHWR-----SPGPGSDQOQERTPLARLP 206
Qy 268 -----ARGMLHLPSCAERRAFLLSLQATVALMMNCGIT-----PI 303
Db 207 AVAADAGRGGLQLPPLPQHVGERDAPPAILCRAEGFLTVPSGHRGPHHELWHYPI 266
Qy 304 IYFFASTHYRKWLGLILKLGSSSSSSSSSTPGKASSETPSITQARGSMFLAEHV 360
Db 267 IYFFASTHYRKWLGLILKLGSSSSSSSSSTPGKASSETPSITQARGSMFLAEHV 323

RESULT 3
US-10-292-798-1294
; Sequence 1294, Application US/10292798
; Publication No. US2003023583A1
; GENERAL INFORMATION:

; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1294
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-292-798-1294

Query Match 47.1%; Score 884; DB 12; Length 323;
Best Local Similarity 65.7%; Pred. No. 4.7e-73;
Matches 195; Conservative 8; Mismatches 36; Indels 58; Gaps 5;
Qy 101 GLCRLTAFVLYTDTYGGVILMACVSDHYPAVCAHMGPRLTAGARLVCAIWTLLV 160
Db 48 GVCALHRLH-----RGVYLMACVSDHYPAVCAHMGPRLTAGARLVCAIWTLLV 101
Qy 161 QTMPELLMPMTKPLVGLKACMEYSMESVGLGLPLMLVAFAGFCGPGVGIILSCYNKTIW 220
Db 102 QTMPELLMPMTKPLVGLKACMEYSMESVGLGLPLMLVAFAGFCGPGVGIILSCYNKTIW 161
Qy 221 KLCs-----TAREDPVTSRGRHWRGCLLTLLMLVAVVCFSPYHLNKKQFM----- 267
Db 162 KLCSTAGRTQAREDPVTSRGRHWR-----SPGPGSDQOQERTPLARLP 206
Qy 268 -----ARGMLHLPSCAERRAFLLSLQATVALMMNCGIT-----PI 303
Db 207 AVAADAGRGGLQLPPLPQHVGERDAPPAILCRAEGFLTVPSGHRGPHHELWHYPI 266
Qy 304 IYFFASTHYRKWLGLILKLGSSSSSSSSSTPGKASSETPSITQARGSMFLAEHV 360
Db 267 IYFFASTHYRKWLGLILKLGSSSSSSSSSTPGKASSETPSITQARGSMFLAEHV 323

RESULT 4
US-10-041-615-109
; Sequence 109, Application US/10041615
; Publication No. US20040014038A1
; GENERAL INFORMATION:
; APPLICANT: Caeman, Stacie J
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Ellerman, Karen
; APPLICANT: Smithson, Glenda
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: No. US20040014038A1el GPCR-Like Proteins and Nucleic Acids Encodi
; FILE REFERENCE: 21402-233-061
; CURRENT APPLICATION NUMBER: US/10/041,615
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: 60/259,552
; PRIOR FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: 60/260,544
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/277,405
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: CuraSeqlist version 0.1
; SEQ ID NO 109
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-041-615-109

Query Match 38.4%; Score 721; DB 12; Length 229;
Best Local Similarity 85.1%; Pred. No. 3e-58;
Matches 143; Conservative 5; Mismatches 12; Indels 8; Gaps 2;

QY 59 SRKINCTGYLVHVAIVDGLFTVALPGRVVCVVGSSWPFKGLCLRTAFVLYTDTYGVV 118
DB 66 SRKINCTGYLVHVAIVDGLFTVALPGRVVCVVGSSWPFKGLCLRTAFVLYTDTYGVV 117
QY 119 YLMACVSDVHPAVVCAHWPGRVLCVVAIVTWTLLVLTQWPLLMPTKPLVGLK 178
DB 118 YLMACVSDVHPAVVCAHWPGRVLCVVAIVTWTLLVLTQWPLLMPTKPLVGLK 177
QY 179 ACNEYSMESVGLPLMVLVAFVAFGCGVGVGILSCYMKITWKLSTA 226
DB 178 ACNEYSMESVGLPLMVLVAFVAFGCGVGVGILSCYMKITWKLSTA 225

RESULT 5
US-09-827-937A-17
; Sequence 17, Application US/09827937A
; Patent No. US2002052043A1
; GENERAL INFORMATION:
; APPLICANT: Li, Yi
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human G-Protein Coupled Receptors
; FILE REFERENCE: 1488.122003
; CURRENT APPLICATION NUMBER: US/09/827, 937A
; CURRENT FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: 06/852,824
; PRIOR FILING DATE: 1997-05-07
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-827-937A-17

Query Match 33.9%; Score 635.5; DB 9; Length 348;
Best Local Similarity 38.5%; Pred. No. 3.6e-50;
Matches 132; Conservative 69; Mismatches 139; Indels 3; Gaps 2;

QY 2 IKLGYPPTPTKTKIKDLNVSQTVKIFLRSFYTALLVFSALGNILALCLTCQKSR 61
DB 1 IQMANNTPPSATPQNDCLYAHHSARIVM-PLHYSLVFIILGLVGNLLALVIVQNRK 59
QY 62 INCTGYLVHVAIVDGLFTVALPGRVVCVVGSSWPFKGLCLRTAFVLYTDTYGVV 121
DB 60 INSTLLSTNLVSDILFTTALPRTIAYANGDWRIGDLCALCRITAFVLYTDTYGVV 119
QY 122 ACVSDVHPAVVCAHWPGRVLCVVAIVTWTLLVLTQWPLLMPTKPLVGLK 181
DB 120 TCLSIDRFIAVHPLRYNKKRIEHAQVCFVWILVFAQTLLPLINPMSKQAEARITC 179
QY 182 EYSSMESVGLPLMVLVAFVAFGCGVGVGILSCYMKITWKLSTAEDPVTSRKGRHW 241
DB 180 EYPNFEETKSLPWTLLGACFGVGVVPLIIILICVQICCKLFRITAKONPLTEKSGV 239
QY 242 CLLTLLMVAIVVCFSPYHLNLIKQFMARGM--LHLPSCAERRAFLLSLQATVLMNMC 299
DB 240 LNTIILIVVFLCTFPHVAIIQHMKLRFNLFCSQHSFQISLHFTVCLMNFNC 299
QY 300 ITPIIYFFASTHYRKLWLLGILKLGSSSSSSSSSTPGKAS 342
DB 300 MDPIYFFACKGYKRVMLKRVQSVSISSAVKSAPEENSREMTQWMIHKS 342

RESULT 6
US-10-041-615-110
; Sequence 110, Application US/10041615
; Publication No. US20040014038A1
; GENERAL INFORMATION:
; APPLICANT: Casman, Stacie J
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Ellerman, Karen
; APPLICANT: Smithson, Glennda
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Padigar, Muralidhara
; TITLE OF INVENTION: No. US20040014038A1 GPCR-Like Proteins and Nucleic Acids Encod
; FILE REFERENCE: 21402-233-061
; CURRENT APPLICATION NUMBER: US/10/041, 615
; CURRENT FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: 60/259,552
; PRIOR FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: 60/260,544
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/277,405
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 110
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-041-615-110

Query Match 33.7%; Score 632.5; DB 12; Length 361;
Best Local Similarity 37.8%; Pred. No. 7.1e-50;
Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;

QY 2 IKLGYPPTPTKTKIKDLNVSQTVKIFLRSFYTALLVFSALGNILALCLTCQKSR 60
DB 3 IQMANNTPPSATPQNDCLYAHHSARIVM-PLHYSLVFIILGLVGNLLALVIVQNRK 61
QY 61 KINCTGYLVHVAIVDGLFTVALPGRVVCVVGSSWPFKGLCLRTAFVLYTDTYGVV 120
DB 62 KINSTLLSTNLVSDILFTTALPRTIAYANGDWRIGDLCALCRITAFVLYTDTYGVV 121
QY 121 YACVSDVHPAVVCAHWPGRVLCVVAIVTWTLLVLTQWPLLMPTKPLVGLK 180
DB 122 TCLSIDRFIAVHPLRYNKKRIEHAQVCFVWILVFAQTLLPLINPMSKQAEARITC 181
QY 181 MEYSMESVGLPLMVLVAFVAFGCGVGVGILSCYMKITWKLSTAEDPVTSRKGRHW 240
DB 182 MEYPNFEETKSLPWTLLGACFGVGVVPLIIILICVQICCKLFRITAKONPLTEKSGV 241
QY 241 GCLLTLLMVAIVVCFSPYHLNLIKQFMARGM--LHLPSCAERRAFLLSLQATVLMNMC 298
DB 242 ALNTIILIVVFLCTFPHVAIIQHMKLRFNLFCSQHSFQISLHFTVCLMNFNC 301
QY 299 GITPIIYFFASTHYRKLWLLGILKLGSSSSSSSSSTPGKAS---SETPSITQARG 352
DB 302 CNDPIYFFACKGYKRVMLKRVQSVSISSAVKSAPEENSREMTQWMIHKS 358

RESULT 7
US-10-222-024-2
; Sequence 2, Application US/10222024
; Publication No. US20030104487A1
; GENERAL INFORMATION:
; APPLICANT: Pfizer Ltd. (GB)
; APPLICANT: Pfizer Inc. (EP except GB, US, JP)
; TITLE OF INVENTION: Neuropeptide receptor and uses thereof
; FILE REFERENCE: PCS22032
; CURRENT APPLICATION NUMBER: US/10/222,024
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: GB 0119920.7
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-222-024-2


```

; PRIOR APPLICATION NUMBER: 60/217,370
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/218,492
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 60/186,810
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: 60/188,064
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: 60/186,457
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: 60/213,861
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/194,344
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/218,337
; PRIOR FILING DATE: 2000-07-14
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 95
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-932-95

Query Match      22.2%; Score 417; DB 11; Length 177;
Best Local Similarity 58.9%; Pred. No. 2e-30;
Matches 93; Conservative 8; Mismatches 17; Indels 40; Gaps 5;

QY 101 GLCRLTAFVLYTDTGGVILMACVSDYHYPANVCANHGPRILRTAGRLVCVAIWTLLV 160
Db 14 GVCLARHL-----RGVILMACVSDYHYPANVCANHGPRILRTAGRLVCVAIWTLLV 67

QY 161 QTPFLLLMPMTKPLVGKLACMEYSSMESVLGLPLMLVLVAFAGFCGPVGIIISCYMKITW 220
Db 68 QTPFLLLMPMTKPLVGKLACMEYSSMESVLGAAPH-----GPGGL---CH-----W 110

QY 221 KLCST-----AREDPVTSRKGRHWRG 241
Db 111 LLWASGDHPVLLYEDHLEAVQHSWENPVTSGKHRRG 148

RESULT 12
US-10-251-385-182
; Sequence 182, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human
; TITLE OF INVENTION: Protein-Coupled
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/10/251,385
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US/09/170,496
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 182
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-251-385-182

Query Match      21.1%; Score 395.5; DB 15; Length 339;
Best Local Similarity 30.5%; Pred. No. 4.1e-26;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;

QY 36 FYTALLVFSALGNILCLTCQSRKINCTGIVLHVLAVSDLLFTVALPGRWCYVLGSS 95
Db 37 FYLLDFILAVGNTLALWLPFIRDKSGTTPANVFLMHLAVADLSCVLVLPTRLVYHFSGNH 96

```

QY 96 WPGKGLCLTAFVLYTDTYGGVYLMACVSDHYPAVCAHWGPRRTAGARLVCAI 155
Db 97 WPFGEIACRLTGLFYLANYASITFTICISADRFIAIHPVKSUKRRPLRYAHLACAF 156
QY 156 TLVLQTMPLLLMPMTKPLVGLACME--YSSMESVGLGLMLVLAFAIGFCGPGVGIILSC 214
Db 157 VVAVAMAPLLVSPQTVQTNHTVCLQYREKAS-----HHAIVSLAVAFTPFITTTVC 211
QY 215 YMKITWKLCTAREDPVTSRKGHRWGCLLTLLMLVAVVCFSPYHLINIKOPMARGMLHL 274
Db 212 YLLIIRSLRQGRVE-----KELKTAVRMIAIHLAFLVCFVPHVNRSVYVLYHRS 266
QY 275 PSACERRAPLLSLQATVALMNMNCGITPIIYFFASTHYRKMLLGIL---KLKG---SSSS 328
Db 267 ASCATORILANRITSLTSLNGALDPIYFFVAEKFRHALCNLLCGKRLKGPSPFEG 326
QY 329 SSSSSSTPGKA 339
Db 327 KTNESLSAKS 337

RESULT 13
US-09-848-889-12
; Sequence 12, Application US/0984889
; Patent No. US2002002555A1
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Guegler, Karl J.
; APPLICANT: Cheng, Muzong
; TITLE OF INVENTION: GPCR DIAGNOSTIC FOR BRAIN CANCER
; FILE REFERENCE: PC-0042 CIP
; CURRENT APPLICATION NUMBER: US/09/848,889
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PERL Program
; SEQ ID NO 12
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US2002002555A1 9992700
US-09-848-889-12

Query Match 20.9%; Score 391.5; DB 9; Length 339;
Best Local Similarity 30.5%; Pred. No. 9.5e-28;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;

QY 36 FYTALLVFSALGNILALCLTCOKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
Db 37 FYLLDFILALVGNLTALWLFIRDHKSGETPANVFLMHLAVADLSCVLPLTRLVYHFSGNH 96
QY 96 WPGKGLCLTAFVLYTDTYGGVYLMACVSDHYPAVCAHWGPRRTAGARLVCAI 155
Db 97 WPFGEIACRLTGLFYLANYASITFTICISADRFIAIHPVKSUKRRPLRYAHLACAF 156
QY 156 TLVLQTMPLLLMPMTKPLVGLACME--YSSMESVGLGLMLVLAFAIGFCGPGVGIILSC 214
Db 157 VVAVAMAPLLVSPQTVQTNHTVCLQYREKAS-----HHAIVSLAVAFTPFITTTVC 211
QY 215 YMKITWKLCTAREDPVTSRKGHRWGCLLTLLMLVAVVCFSPYHLINIKOPMARGMLHL 274
Db 212 YLLIIRSLRQGRVE-----KELKTAVRMIAIHLAFLVCFVPHVNRSVYVLYHRS 266
QY 275 PSACERRAPLLSLQATVALMNMNCGITPIIYFFASTHYRKMLLGIL---KLKG---SSSS 328
Db 267 ASCATORILANRITSLTSLNGALDPIYFFVAEKFRHALCNLLCGKRLKGPSPFEG 326
QY 329 SSSSSSTPGKA 339
Db 327 KTNESLSAKS 337

RESULT 14
US-09-788-133-2
; Sequence 2, Application US/09788133
; Patent No. US20020052001A1
; GENERAL INFORMATION:
; APPLICANT: GLAXO GROUP LTD
; TITLE OF INVENTION: ASSAY
; FILE REFERENCE: P79011
; CURRENT APPLICATION NUMBER: US/09/788,133
; CURRENT FILING DATE: 2001-02-20
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-788-133-2

Query Match 20.9%; Score 391.5; DB 9; Length 339;
Best Local Similarity 30.5%; Pred. No. 9.5e-28;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;

QY 36 FYTALLVFSALGNILALCLTCOKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
Db 37 FYLLDFILALVGNLTALWLFIRDHKSGETPANVFLMHLAVADLSCVLPLTRLVYHFSGNH 96
QY 96 WPGKGLCLTAFVLYTDTYGGVYLMACVSDHYPAVCAHWGPRRTAGARLVCAI 155
Db 97 WPFGEIACRLTGLFYLANYASITFTICISADRFIAIHPVKSUKRRPLRYAHLACAF 156
QY 156 TLVLQTMPLLLMPMTKPLVGLACME--YSSMESVGLGLMLVLAFAIGFCGPGVGIILSC 214
Db 157 VVAVAMAPLLVSPQTVQTNHTVCLQYREKAS-----HHAIVSLAVAFTPFITTTVC 211
QY 215 YMKITWKLCTAREDPVTSRKGHRWGCLLTLLMLVAVVCFSPYHLINIKOPMARGMLHL 274
Db 212 YLLIIRSLRQGRVE-----KELKTAVRMIAIHLAFLVCFVPHVNRSVYVLYHRS 266
QY 275 PSACERRAPLLSLQATVALMNMNCGITPIIYFFASTHYRKMLLGIL---KLKG---SSSS 328
Db 267 ASCATORILANRITSLTSLNGALDPIYFFVAEKFRHALCNLLCGKRLKGPSPFEG 326
QY 329 SSSSSSTPGKA 339
Db 327 KTNESLSAKS 337

RESULT 15
US-09-828-478-4
; Sequence 4, Application US/09828478
; Patent No. US20020155528A1
; GENERAL INFORMATION:
; APPLICANT: Xiao, Yonghong
; TITLE OF INVENTION: Regulation of Human CytLT2-Like GPCR
; FILE REFERENCE: 04974.00458
; CURRENT APPLICATION NUMBER: US/09/828,478
; CURRENT FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: 60/195,196
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/254,876
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-828-478-4

Query Match 20.9%; Score 391.5; DB 10; Length 339;
Best Local Similarity 30.5%; Pred. No. 9.5e-28;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;

Search completed: February 19, 2004, 19:36:22
Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 19:30:15 ; Search time 20 Seconds
(without alignments)

1731.036 Million cell updates/sec

Title: US-10-041-615-34

Perfect score: 1876

Sequence: 1 MIKGPYPTPTKTKTKKD.....SETPSITQAGSMFLAEHV 360

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	632.5	33.7	361	B45680	G protein-coupled
2	364.5	19.4	427	S17148	alpha-thrombin rec
3	339.5	18.1	425	A37912	thrombin receptor
4	339.5	18.1	432	A34448	thrombin receptor
5	332.5	17.7	420	I51667	thrombin receptor
6	329.5	17.5	328	JC4800	P2Y6 receptor - hu
7	324.5	17.3	355	JC5067	G protein-coupled
8	322.5	17.2	360	A57160	chemokine (C-C) re
9	315.5	16.8	371	JC5498	G protein-coupled
10	308.5	16.4	328	I55450	G protein-coupled
11	308.5	16.4	373	A47556	ATP receptor P2u -
12	307.5	16.4	362	A3714	G protein-coupled
13	307.5	16.4	355	A45177	chemokine (C-C) re
14	306.5	16.3	360	A53611	interleukin-8 rece
15	302.5	16.1	380	I38435	angiotensin recept
16	300.5	16.0	371	JC5796	probable chemotr
17	300.5	16.0	367	JC0349	interferon-inducib
18	298.5	15.9	358	A53752	interleukin-8 rece
19	298.5	15.9	359	A49241	interleukin-8 rece
20	297.5	15.9	308	I50241	G protein-coupled
21	297.5	15.8	355	JQ1231	heptahelical P2Y5-
22	296.5	15.8	370	JC5549	G protein-coupled
23	293.5	15.6	362	A30341	probable G protein
24	293.5	15.6	395	A40685	intronic 17 puriner
25	291.5	15.5	344	T09508	P-2U nucleotide re
26	290.5	15.5	375	A54946	probable G protein
27	290.5	15.5	428	S30508	MIP-1 alpha recept
28	289.5	15.4	359	I49341	interleukin-8 rece
29	287.5	15.3	350	A39445	

RESULT 2

S17148

alpha-thrombin receptor - Chinese hamster

ALIGNMENTS

RESULT 1

B45680

G protein-coupled peptide receptor EBI 2 - human

C:Species: Homo sapiens (man)

C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000

C:Accession: B45680

J. Birkenbach, M.; Josefsen, K.; Yalamanchili, R.; Lenoir, G.; Kieff, E.

J. Virol. 67, 2209-2220, 1993

A:Title: Epstein-Barr virus-induced genes: first lymphocyte-specific G protein-coupled p

A:Reference number: A45680; MUID:93188173; PMID:8383238

A:Accession: B45680

A>Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-361 <BIR>

A:Cross-references: GB:I08177; NID:G292056; PIDN:AAA35924.1; PID:G292057

A:Experimental source: B-lymphocytes

A:Note: sequence extracted from NCBI backbone (NCBIN:127096, NCBIP:127097)

C:Superfamily: ATP receptor P2u

C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 33.7%; Score 632.5; DB 2; Length 361;

Best Local Similarity 37.8%; Pred. No. 5.5e-48;

Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;

QY	2	IKLGPYPTPTKTI-KTKIKDLNVESQTVKIFLRSFYTALLVFSALGNILALCLTCQKSR	60
DB	3	IQVANNFTPPSATPQGNDCDLVAHSTARIUM-PLHYSLVFFIIGLVGNLLALVVIQNK	61
QY	61	KINCTGIYLVHLAVSDLLFTVALPGRWCVLGSWPFQKGLCRLLTAFLVYTDYTGCVL	120
DB	62	KINSTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITAFVINTYAGVNF	121
QY	121	MACVSDVHPVAVCAHWGPRLRTAGRLVCVAITWLVLTQTMPLLLMPTKPLVGKLC	180
DB	122	MTCLSDIRFIADVHPRLRYNKKRIEAKGVCFVWLVFAQLPLLINPMSQEAERITC	181
QY	181	MEYSSMESVGLPLMVLVAFIAGCPGVGIIISCYMKITWKLCTAREDPVTSRKQRHR	240
DB	182	MEYPNFEETKSLFWLLGACFGVLPILIIICYSQICCKLFRYAKQNPLEKSGVKK	241
QY	241	GCULTILMLVAVVVCSPHLNKKQPMRGM-LHLPSCAERRAFLLSLQATVALNMNC	298
DB	242	ALNTIILIIIVVLCFTFVVAIQIMIKKLFARFNLECSQRHSFOISLHFTVCLMNFNC	301
QY	299	GITPIIYFFASTHYRWKLGLIKLKGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS	352
DB	302	CMDPFYIFACKGYKKVMRLKQVSVSISSAVKSAPEENSREMTETQWTHSKSS	358

C;Species: Cricetulus griseus (Chinese hamster)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C;Accession: S17148
R;Rasmussen, U.B.; Vouret-Craviari, V.; Jallat, S.; Schlesinger, Y.; Pages, G.; Pavitan
FEBS Lett. 288, 123-128, 1991
A;Title: cDNA cloning and expression of a hamster alpha-thrombin receptor coupled to Ca
A;Reference number: S17148; MUID:91348247; PMID:1652467
A;Accession: S17148
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-427 <RAS>
A;Cross-references: EMBL:X61958; NID:g940495; PIDN:CAA43957.1; PID:g49538
C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match 19.4%; Score 364.5; DB 2; Length 427;
Best Local Similarity 30.7%; Pred. No. 2.3e-24;
Matches 96; Conservative 58; Mismatches 146; Indels 13; Gaps 6;

Qy 25 SQTWVIFLRSFYTALLVPSALGNILALCLTCQSKRNCTGIVLVHLAVSDILFTVALP 84
Db 100 SFMLRFLFIPSV-YTFVFWLSPLNLAIAVFLVLMKVKPAVVMHLAVDLFVSVLP 158

Qy 85 GRVVCYVVGSSWPGKGLCLRTAFVLTDTYGGVYLMAVCVSDHYPAVCAHWGPRLRTA 144
Db 159 LKISYFSGSDWQFGSGMCRFATAAFYCNMYASIMLTMTVISIDRFLAVVPIQSLSWRTL 218

Qy 145 GRARLVCVAIWTLLVLTQMPILLMPMTK--PLVGLKACMEYSMESVGLGLPLM-VLF 201
Db 219 GRANFTCLVIMWAIMGVWPLLEKEQTRVPGNITTCDDVLAETLQGGYSYFSAFA 278

Qy 202 IGFCGPGVIGILSCYMKITWKLCSTAREDPVTSRGRHWRGCLLTLLMLAVVVCFSFYHL 261
Db 279 VFPLVPLIISTICYMSIIRCLSSSS-----VANRSKSRALFLSAAVFCVIFVCGFP--T 331

Qy 262 NIKQFMARGMLHPSCAERAFILSLQATVALNMNCGITPIIYFFASTHYRKMILGILK 321
Db 332 NVLLIMHYLLSDSPATEKAYFAYLL--CVCVSSVCCIDPLIYYASSSECQRTLYGILC 389

Qy 322 LKGSSSSSSSSSSS 334
Db 390 CKESDDNSYNT 402

RESULT 3
A37912
thrombin receptor precursor - human
C;Species: Homo sapiens (man)
C;Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 05-Nov-1999
C;Accession: A37912
R;Vu, T.K.H.; Hung, D.T.; Wheaton, V.I.; Coughlin, S.R.
Cell 64, 1057-1068, 1991
A;Title: Molecular cloning of a functional thrombin receptor reveals a novel proteolytic
A;Reference number: A37912; MUID:91168254; PMID:1672265
A;Accession: A37912
A;Molecule type: mRNA
A;Residues: 1-425 <VUA>
A;Cross-references: GS:M62424; NID:g339676; PIDN:AAA36743.1; PID:g339677
C;Genetics:
A;Gene: GDB:F2R
A;Cross-references: GDB:127737; OMIM:187930
A;Map position: 5q13-5q13
C;Keywords: G protein-coupled receptor; glycoprotein; transmembrane protein
F11-26/Domain: signal sequence #status predicted <SIG>
F127-425/Product: thrombin receptor #status predicted <MAT>

Query Match 18.1%; Score 339.5; DB 2; Length 425;
Best Local Similarity 29.1%; Pred. No. 3.7e-22;
Matches 98; Conservative 58; Mismatches 160; Indels 21; Gaps 7;

Qy 25 SQTWVIFLRSFYTALLVPSALGNILALCLTCQSKRNCTGIVLVHLAVSDILFTVALP 84
Db 98 SSWLTLFVPSV-YTGVFVVSPLNLAIAVFLVLMKVKPAVVMHLATADLVFVSVLP 156

Qy 85 GRVVCYVVGSSWPGKGLCLRTAFVLTDTYGGVYLMAVCVSDHYPAVCAHWGPRLRTA 144
Db 157 PKISYFSGSDWQFGSGMCRFATAAFYCNMYASIMLTMTVISIDRFLAVVPMOSLSWRTL 216

Qy 145 GRARLVCVAIWTLLVLTQMPILLMPMTKPLVGLKACMEYSMESVGLGLPLM-----VLV 198
Db 217 GRASFTCLAIWALAIAGVPLVLEQTIQVPG-----LNITCHDVNETLLEGGYAYYFS 272

Qy 199 AF-AIGFCGPGVIGILSCYMKITWKLCSTAREDPVTSRGRHWRGCLLTLLMLAVVVCFS 257
Db 273 AFSAVFFVPLIISTVCVSIIRCLSSA-----VANRSKSRALFLSAAVFCIFCFG 327

Qy 258 PYHLNMQFMARGMLHPSCAERAFILSLQATVALNMNCGITPIIYFFASTHYRKMIL 317
Db 328 P--TNVLLIAHYSFLSHTSITTEAAYFAYLL--CVCVSSISCIDPLIYYASSSECQVYV 383

Qy 318 GILKLGSS 354
Db 384 SILCKESSDPSYNSQGLMASKMDTCSSNLNNSIY 420

RESULT 4
A43448
thrombin receptor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C;Accession: A43448
R;Zhong, C.; Hayzer, D.J.; Corson, M.A.; Runge, M.S.
J. Biol. Chem. 267, 18975-18979, 1992
A;Title: Molecular cloning of the rat vascular smooth muscle thrombin receptor. Evidence
A;Reference number: A43448; MUID:92381002; PMID:1324917
A;Accession: A43448
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-432 <ZHO>
A;Cross-references: GS:M81642; NID:g207465; PIDN:AAA42274.1; PID:g207466
A;Experimental source: RASM aortic smooth muscle cells
A;Note: sequence extracted from NCBI backbone (NCBIN:111973, NCBIPI:111974)
C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match 18.1%; Score 339; DB 2; Length 432;
Best Local Similarity 29.0%; Pred. No. 4.1e-22;
Matches 98; Conservative 61; Mismatches 153; Indels 26; Gaps 9;

Qy 6 PYFTPTTKIKIKIDLN--VESQVVKIFLRLFTTALLVFSALGNILALCLTCQSKRN 63
Db 87 PMPPPPPTIS--EDASGYLTSPMLTFLPSV-YTFVFIIVSLPLNLAIAVFPVPMKYK 142

Qy 64 CTGIYLVHLAVSDILFTVALPGRVVCYVVGSSWPGKGLCLRTAFVLTDTYGGVYLMA 123
Db 143 PAVVTMLHLAVDLFVSVLPFKISYFSGTDWQFGSGMCRFATAACYNMYASIMLTMTV 202

Qy 124 VSDHYPAVCAHWGPRLRTAGARLVCVAIWTLLVLTQMPILLMPMTKPLVGLKACMEY 183
Db 203 ISIDRFLAVVPIQSLSWRTLGRANFTCVIWMVMAIMGVWPLLEKEQTIQVPG-----LNI 258

Qy 184 SSMESVGLGLPLM-----VLVAF-AIGFCGPGVIGILSCYMKITWKLCSTAREDPVTSRKG 236
Db 259 ITCHDVNETLHGYSYFSAFSAIFLVLPLIISTVCYTSIIRCLSSA-----VANRS 313

Qy 237 RHWRCCLLTLLMLAVVVCFSFYHLNMQFMARGMLHPSCAERAFILSLQATVALNMN 296
Db 314 KKSRAFLSAAVFCIFVCGFTNVLL---IVHLLSDSPGTETAYFAYLLCVCVTSVA 370

Qy 297 NCGITPIIYFFASTHYRKMILGILKLGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 334
Db 371 SC-IDPLIYYASSSECQKHLVILCCRESSDSNSCNSIT 407

RESULT 5
I51667
thrombin receptor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)

C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C/Accession: J151667
R/Gersten, R.E.; Chen, J.; Ishii, M.; Nanevitz, T.; Turck, C.W.; Vu, T.H.; Q
Nature 368, 648-651, 1994
A>Title: Thrombin receptor's specificity for agonist peptide is determined by its extrac
A/Reference number: J151667; MUID:94195429; PMID:8145052
A/Accession: J151667
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-420 <GER>
A/Cross-references: EMBL:U09632; NID:9495197; PIDN:AAA18498.1; PID:g495198

Query Match 17.7%; Score 332.5; DB 2; Length 420;
Best Local Similarity 26.6%; Pred. No. 1.5e-21;
Matches 92; Conservative 74; Mismatches 155; Indels 25; Gaps 8;
QY 16 TKIKDLNVEQTKIFLRSFYTALLVFSALGNILALCLTCQSKRKNCTGTYLVLHAYS 75
DB 88 TKEAQVLSQWLTKEFVPSL-YTVFVIGLPLNLAIIIFLFKMKRKAAPVVMNLNLAIA 146
QY 76 DLFTVALPGRVVCVVLGSSWPGKGLCLRTAFVLTDTYGGVYLMACVSDHYPAVUCA 135
DB 147 DVFVSVLPFKIAHYHSGNDWLPFGMCRIVTAIFYCNMYCVSLLIASISVDRFLAVVYP 206
QY 136 HWGPRLRTAGRALVCAIVAIWTLVLLQTMPLLMPTKPLVGLACMEYSMSVGLPLM 195
DB 207 MHSLSWETMSRAYMACSFILWISASTIPLLVTEQIQ-----KIPRLDITTCHEVDLKDOL 262
QY 196 --VLVAPAGFC-----GPVGIILSCYMKITWKLCTAREDPVTSRKGHRGCLLTLM 248
DB 263 KDFVYVYFSCFLFFFPVPIITTCVIGIRSLSSSIENS-----KTRALPLAVV 317
QY 249 LVAVVCFSPYHLNFKQFARGMLHLPSCAERRAFLLSLQATVALMNMCGTPIIYFPA 308
DB 318 LCVFIIICFGTNNV---LFLTH---YLQEAHEFLYFAYILSACVG--SVSCCLDPLIYYA 369
QY 309 STYRWKLLGILKLGSSSSSSSSSTPGKASSETSIITQARSMF 354
DB 370 SSQCQRYLSLLCKRVSEPGSGTGLMSTAMKNDNCSTNAKSSIV 415

RESULT 6

JC4800
P2Y6 receptor - human
C/Species: Homo sapiens (man)
C/Date: 15-Oct-1995 #sequence_revision 16-Aug-1996 #text_change 17-Nov-2000
C/Accession: JC4800; G02514
R/Communi, D.; Parmentier, M.; Bosyns, J.M.
Biochem. Biophys. Res. Commun. 222, 303-308, 1996
A>Title: Cloning, functional expression and tissue distribution of the human P2Y6 recep
A/Reference number: JC4800; MUID:96222498; PMID:8670200
A/Accession: JC4800
A/Molecule type: mRNA
A/Residues: 1-328 <COM>
A/Cross-references: EMBL:X97058
A/Experimental source: placenta
R/Hammet, F.; Southey, M.C.; Somers, G.R.; Hutchins, A.M.; Venter, D.J.
submitted to the EMBL Data Library, March 1996
A/Reference number: H01373
A/Accession: G02514
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-328 <HAM>
A/Cross-references: EMBL:U52464; NID:gl407632; PIDN:AAB03572.1; PID:gl407633
C/Genetics:
A/Gene: P2Y6
C/Supfamily: ATP receptor P2u
A/Keywords: Glycoprotein; placenta; receptor; transmembrane protein
P/26-52/Domain: transmembrane #status predicted <TM1>
P/63-86/Domain: transmembrane #status predicted <TM2>
F/104-122/Domain: transmembrane #status predicted <TM3>
F/143-167/Domain: transmembrane #status predicted <TM4>
F/193-216/Domain: transmembrane #status predicted <TM5>

F/241-264/Domain: transmembrane #status predicted <TM6>
F/283-305/Domain: transmembrane #status predicted <TM7>
F/5,173/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 17.5%; Score 329; DB 2; Length 328;
Best Local Similarity 28.0%; Pred. No. 2.4e-21;
Matches 89; Conservative 54; Mismatches 141; Indels 34; Gaps 11;

QY 10 PPTKIKTKIDLVNESQTKIFLRSFYTALLVFSALGNILALCLTCQSKRKNCTGIYL 69
DB 14 PPTTC-----VYRENFKQLLPPVYSAVLAAGLPLNICVITQICTSRALRTAVYT 65
QY 70 VHLAVSLLFTVALPGRVVCVVLGSSWPGKGLCLRTAFVLTDTYGGVYLMACVSDHY 129
DB 66 LNLALADLLYACSLPLLIYNAQGDHFFGDFACRLVFLFYANLHSLILELTICISQRY 125
QY 130 PAVCAHWGPRLRAGR--ARLVCAVIAWTLVLLQTMPLLMPTKPLVGLACMEYS--- 184
DB 126 LG-ICHPLAPWHKRGRRRAWLVCAVWLAVTTQCLPTAIPAAATGIQRNRTVCYDLSPPA 184
QY 195 --SMESVLGLPLVLAFAIGFCPVGIILSCYMKITWKLCTAREDPVTSRKGHRGCLLTLM 237
DB 185 LATHYMPYGMALT-----IGLFFFAALLACYCLLACRLC---RQDGPAPVQAQERAGK 236
QY 238 HWRGCLLTLLMLVAVVCFSPYHLNFKQFMA-RGMLHLPSCAERRAFLLSLQATVALMNM 296
DB 237 AAR---NAVVAFAAISFLPFHITKTAYLAVRSTPGVP-CTVLEAFAAAAYKGRPFASA 292
QY 297 NCGITPIIYFPASTHYRK 314
DB 293 NSVLDPLIFYFTQKKFER 310

RESULT 7

JC5067
G protein-coupled receptor CRK-L1 - human
N/Alternate names: chemokine receptor-like protein TER1; GPR-CY6
C/Species: Homo sapiens (man)
C/Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 21-Jul-2000
C/Accession: JC5067; G02776; G02387
R/Zaballos, A.; Varona, R.; Gutierrez, J.; Lind, P.; Marquez, G.
Biochem. Biophys. Res. Commun. 227, 846-853, 1996
A>Title: Molecular cloning and RNA expression of two new human chemokine receptor-like
A/Reference number: JC5067; MUID:97040707; PMID:8886020
A/Accession: JC5067
A/Molecule type: DNA
A/Residues: 1-355 <ZAB>
A/Cross-references: EMBL:Z79782; NID:gl668735; PIDN:CAB02142.1; PID:gl668736
R/Napolitano, M.; Zingoni, A.; Bernardini, G.; Spinetti, G.; Rocchi, M.; Santoni, A.
submitted to the EMBL Data Library, June 1996
A/Reference number: H01714
A/Accession: G02776
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-355 <NAP>
A/Cross-references: EMBL:U62556; NID:gl468978; PID:gl468979
R/Bonner, T.I.
submitted to the EMBL Data Library, January 1996
A/Reference number: H01154
A/Accession: G02387
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-355 <BON>
A/Cross-references: EMBL:U45983; NID:gl245056; PID:gl245057
C/Comment: This protein belongs to the family of beta chemokine receptors.
C/Genetics:
A/Gene: GDB:CMKBR8; CMKBR12; TER1; CRK-L1
A/Cross-references: GDB:6053733; OMIM:601834
A/Map position: 3p21-3p21
C/Supfamily: vertebrate rhodopsin
C/Keywords: G protein-coupled receptor; transmembrane protein
F/36-63/Domain: transmembrane #status predicted <TM1>
F/73-94/Domain: transmembrane #status predicted <TM2>

F:108-129/Domain: transmembrane #status predicted <TM3>
F:147-171/Domain: transmembrane #status predicted <TM4>
F:200-222/Domain: transmembrane #status predicted <TM5>
F:239-260/Domain: transmembrane #status predicted <TM6>
F:281-304/Domain: transmembrane #status predicted <TM7>

Query Match 17.3%; Score 324; DB 2; Length 355;
Best Local Similarity 26.4%; Pred. No. 7.1e-21;
Matches 89; Conservative 65; Mismatches 145; Indels 38; Gaps 8;

QY 26 QTVKIFRSFYTALLVFSALGN--ILALCLTCQSKINCTGIYLVHLAVSLLFTVAL 83
DB 31 QINCKLLAVFYCLLFVFSLLGNSLVILVVCKLRSI--TDVLLNLALSLLVFPFSF 88
QY 84 PGRVVCYVYGSSWPGKGLCLRTAFVLYTDYGGVYLMACVSDHYPAVCAHWGRLR 143
DB 89 PFQT--YLLDQWFGTVMCKWSGFGYIGFYSSMFTITLMSVDRLAVAVHAYVALK 146
QY 144 AGRARLVCAVLTLLVLOTMPLELLMPMTKPLVGLKACMEYSSMESVLGLPLMVLVAF 203
DB 147 IRMGTTLCIAVLTALVMTATIIPLLFPYQVASEDGVLCYSFYNQQT---LWKI 203
QY 204 FCG--PVGIIISCYMKITWL--CSTAREDPVTSRKRHWGCGLLFTLLMLVAVV 258
DB 204 ILGLLIPTIRMFVCYIKILHQLKQCNHKT-----KARLVLIIVVIA 253
QY 259 YHMAIKQFMARGMLHLPSCAERRAFLLSLQATVALMNNCGITPIIYFFASTHYR 316
DB 254 FNVVLTSLTSLHSMHLDGCSIQQLTVAHYTEIISFTHCQVNPVIYAFVGRFK 313
QY 317 -----LGILKLGSSSSSSSSSSSPGKASS 341
DB 314 IFQKSCSOFNYLGRQMPRESCXSSSCQSHSSSSSS 350

RESULT 8

A:57160
chemokine (C-C) receptor 4 - human
N:Alternate names: C-C CKR-4
C:Species: Homo sapiens (man)
C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000
C:Accession: A57160
R:Power, C.A.; Meyer, A.; Nemeth, K.; Bacon, K.B.; Hoogewarf, A.J.; Proudfoot, A.E.I.; W
J. Biol. Chem. 270, 19495-19500, 1995
A:Title: Molecular cloning and functional expression of a novel CC chemokine receptor c
A:Reference number: A57160; MUID:95370289; PMID:7642634
A:Accession: A57160
A:Status: Preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-360 <POM>
A:Cross-references: GB:X85740; NID:g1370103; PIDN:CAAS9743.1; PID:g971452
A:Note: source clone K5-5
C:Genetics:
A:Gene: GDB:CMK3R4
A:Cross-references: GDB:677463
A:Map position: 3p21-3p21
C:Keywords: G protein-coupled receptor; glycoprotein; phosphoprotein; transmembrane prot
F:40-65/Domain: transmembrane #status predicted <TM1>
F:76-97/Domain: transmembrane #status predicted <TM2>
F:112-133/Domain: transmembrane #status predicted <TM3>
F:151-175/Domain: transmembrane #status predicted <TM4>
F:208-226/Domain: transmembrane #status predicted <TM5>
F:243-264/Domain: transmembrane #status predicted <TM6>
F:291-308/Domain: transmembrane #status predicted <TM7>
F:29-276/110-187/Disulfide bonds: #status predicted
F:72,350/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predict
F:145/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predicted
F:183,194/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:321/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted

Query Match 17.2%; Score 322.5; DB 2; Length 360;
Best Local Similarity 28.3%; Pred. No. 9.7e-21;

Matches 93; Conservative 58; Mismatches 131; Indels 47; Gaps 9;
QY 29 KIFLASFYTALLVFSALGNILALCLTCQSKINCTGIYLVHLAVSLLFTVALP--GR 86
DB 39 ELFLPPL--YSLVFFVGLGNSVVLVLFKYKRLASMTDVLNLAISDLFFVFSLPFWG- 96
QY 87 VVCYVILGSSWPGKGLCLRTAFVLYTDYGGVYLMACVSDHYPAVCAHWGRLR 146
DB 97 ---YAAQDWFGGLCKNISMWYLVGFYSGIFFVLMISIDRYLAIVHAVFSRLR 153
QY 147 ARLVCVAIWTLLVLOTMPLELLMPMTKPLVGLKACMEYSSMESVLGLPLMVLVAF 206
DB 154 GVITSLATWSAVPASLPGLFSTCYTERNTYCKTKYSLNSTTWKVLSSLEINILGLV 213
QY 207 PVGIILSCMKIT--TWKLCSTAREDPVTSRKRHWGCGLLFTLLMLVAVV 260
DB 214 PLIGIMLCFYSMIIRTLQHCNEKKNAVK-----MIFAVVVLFLGFW 258
QY 261 LNIKQFMAR--GMLHLPSCAERRAFLLSLQATVALMNNCGITPIIYFFASTHYR 317
DB 259 -NIVLFLETLVEVLQDCTFERYLDYAIQATETLAFVHCLNPIIYFFLGKFK 317
QY 318 -----GILKLGSSSSSSSS 331
DB 318 LFKTCRGLFVLQCYCGLLIQYSADTPSS 346

RESULT 9

G protein-coupled receptor DEZ - mouse
C:Species: Mus musculus (house mouse)
C:Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 19-May-2000
C:Accession: JCS498
R:Wethner, A.; Hermey, G.; Schinke, B.; Hermans-Borgmeyer, I.
Biochem. Biophys. Res. Commun. 233, 336-342, 1997
A:Title: A novel G protein-coupled receptor with homology to neurotrophin and chemotr
A:Reference number: JCS498; MUID:97289630; PMID:9144535
A:Contents: Brain
A:Accession: JCS498
A:Molecule type: mRNA
A:Residues: 1-371 <MET>
A:Cross-references: GB:U79525; NID:g1732346; PIDN:AAB53789.1; PID:g1732347
C:Comment: This protein is involved in the bone metabolism.
C:Superfamily: vertebrate rhodopsin
F:110-187/Disulfide bonds: #status predicted

Query Match 16.8%; Score 315.5; DB 2; Length 371;
Best Local Similarity 25.4%; Pred. No. 4.1e-20;
Matches 89; Conservative 67; Mismatches 154; Indels 41; Gaps 9;
QY 24 ESQTVKIFRSFYTALLVFSALGNILALCLTCQSKINCTGIYLVHLAVSLLFTVAL 83
DB 33 EAKVAFVFL--VVIYSLVCFPLGGLGVLIVATFKNKKTNT--VMFVNLAVADFL 90
QY 84 PGRVVCYVILGSSWPGKGLCLRTAFVLYTDYGGVYLMACVSDHYPAVCAHWGRLR 143
DB 91 PXHITVAANDYHWVFGKAMCKISNLLSHNMTYSVFLLTVISFDRICISVLLPWSQNHRS 150
QY 144 AGRARLVCAVLTLLVLOTMPLELLMPMTKPLVGLKACMEYSSMESVLGLP----- 193
DB 151 IRLAYMTCSAVVWLAFPLFSLPSLVRDNTIHHGKTKCFNFSLAAPESPHFAHQVST 210
QY 194 -----LMVLVAFAGFCGPGVGIILSCYMKITWKLCTSTAREDPVTSRKRHWGRLR 247
DB 211 GYSHRVAVTVTRFLCGLFIPVFIITACVUTIVFKL-----QRNLAKNKKPKIIT 262
QY 248 MLVAVVVCSPYH-LNIKQFMARGMLH---LPSCAERRAFLLSLQATVALMNNCGITPI 303
DB 263 IITTFGLPCPYHTLLLE-----LHHTAVPS-----SVFSLGLPLATAVANSCM 312
QY 304 IYFFASTHYRKKLLGLL-LKLGSSSSSSSSSSSPGKASSSETPSITQAGSM 353
DB 313 LYVFMGHDFRKFVKVAFSLANALSBDTGFPSSYPSPHSRFTKMSLNEKASV 363

Db 28 NEDFKVLLPVSYGVVGVGLCLNVALYIFLCLRLKTNASTTTFMFLAVSDSLYASLP 87
QY 85 GRVVCVVLGSSWPFKGLCLRLTAFLVLYTDTYGGVYLMACVSDHYPAV-----CAHWGPR 140
Db 88 LLVYVYARGDHPFSTVLCVLFVLYTNLYCSILFELTCLSVHRCGLVLRPLHSLRW-- 145
QY 141 LRTAGARLVCAVAITWLVLLQTMPLMLMPKFLVGLKACMEYSSMESVGLGLPLMVLVAF 200
Db 146 --EARYARRVAAVVWVVLACQAPVLYFVTTSVRGTRITCHDTSARELSFHFVAYSSVML 203
QY 201 AIGFCGPVGIILSCYMKITWKLCTAREDPVTSRKGHRWGCLLTLMLVAV-VVCFSPY 259
Db 204 GLLFVAVPFSVILVCVYLMARLLKPAY--GTTGGLPRAKRSVTRTALVLAVALCLFLPF 261
QY 260 HLNKQFMARGMLHPLSCAERRAFLLSLQATVALMNMCGITPIIYFFASTHYRWKL--- 316
Db 262 HVTRTLIYFSRDLT--SCHTNAINMAYKITRPLAGANSLCDPLVFLAGQRLVRPADA 320
QY 317 -----LGILK-----LKGSSSSSSSSSSSTPGKASSSTPSI 346
Db 321 KPTEPTPSPQARKGLHRPNRTVRKDLVSDDSRRTSTP--AGSETKDI 371

RESULT 12
A39714
G protein-coupled receptor RDC1 - human
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 24-Nov-1999
C:Accession: A39714
R:Sreedharan, S.P.; Robichon, A.; Peterson, K.E.; Goetzl, E.J.
Proc. Natl. Acad. Sci. U.S.A. 88, 4986-4990, 1991
A:Title: Cloning and expression of the human vasoactive intestinal peptide receptor.
A:Reference number: A39714; PMID:91271317; PMID:1675791
A:Accession: A39714
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-362 <SRE>
A:Cross-references: GB:M64749; NID:G292418; PID:G292419
C:Comment: The authors have revised the name and functional description of this protein
C:Superfamily: vertebrate rhodopsin
C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 16.4%; Score 307.5; DB 2; Length 362;
Best Local Similarity 24.2%; Pred. No. 26-19;
Matches 80; Conservative 74; Mismatches 148; Indels 29; Gaps 6;

QY 22 NVESQTVKIFLSLRYTALLVFSALGNILALCLTCQSKRNCTGTYLVHLAVSDILFTV 81
Db 36 NMPKSVLYTSLFYIFIFVIGMTANSVVVWNIQAKTGYDTHCYILNLAIDLWVVL 95
QY 82 ALPGRVVCVVLGSSWPFKGLCLRLTAFLVLYTDTYGGVYLMACVSDHYPAVCAHWGPR 141
Db 96 TIPVWVSVLVQHNQWPMGELTCKVTHLIFPSINLPSGIFPLTCMSVDVRLSYFTNTPSS 155
QY 142 RTAGARLVCAVAITWLVLLQTMPLMLMPKFLVGLKACM-----EYSSMESVGLGLPLM 195
Db 156 RKMVRRVVCVILWLLAFVCLVLPDYVLYLKTVTSASNNETVCHRSFYPEHSIKEWLIG---M 212
QY 196 VLVAFAIGFCGPGVIGILSCYMKITWKLCTAREDPVTSRKGHRWGCLLTLMLVAVVVC 255
Db 213 ELVSVVLGPAVFPFIIAVFPYFLARAIASASSDQEKHSRK-----IIPSVVVVFLVC 264
QY 256 FSPYHLNITQFMARGMLHLP--SCARRAFLLSLQATVALMNMCGITPIIYFFASTHYRK 314
Db 265 WLPHYHVALLDIFSLIHLIYIPFTLRLEHALFTALHTVQCLSLVHCCVNPVLSFINRY 324
QY 315 WLL-----GILKLGSSSSSSSSSSSSSS 334
Db 325 ELMKAFIPKYSAKTGILKLDASRVSETEYS 355

RESULT 13
A45177

Db 28 NEDFKVLLPVSYGVVGVGLCLNVALYIFLCLRLKTNASTTTFMFLAVSDSLYASLP 87
QY 85 GRVVCVVLGSSWPFKGLCLRLTAFLVLYTDTYGGVYLMACVSDHYPAV-----CAHWGPR 140
Db 88 LLVYVYARGDHPFSTVLCVLFVLYTNLYCSILFELTCLSVHRCGLVLRPLHSLRW-- 145
QY 141 LRTAGARLVCAVAITWLVLLQTMPLMLMPKFLVGLKACMEYSSMESVGLGLPLMVLVAF 200
Db 146 --EARYARRVAAVVWVVLACQAPVLYFVTTSVRGTRITCHDTSARELSFHFVAYSSVML 203
QY 201 AIGFCGPVGIILSCYMKITWKLCTAREDPVTSRKGHRWGCLLTLMLVAV-VVCFSPY 259
Db 204 GLLFVAVPFSVILVCVYLMARLLKPAY--GTTGGLPRAKRSVTRTALVLAVALCLFLPF 261
QY 260 HLNKQFMARGMLHPLSCAERRAFLLSLQATVALMNMCGITPIIYFFASTHYRWKL--- 316
Db 262 HVTRTLIYFSRDLT--SCHTNAINMAYKITRPLAGANSLCDPLVFLAGQRLVRPADA 320
QY 317 -----LGILK-----LKGSSSSSSSSSSSTPGKASSSTPSI 346
Db 321 KPTEPTPSPQARKGLHRPNRTVRKDLVSDDSRRTSTP--AGSETKDI 371

RESULT 11
A47556
ATP receptor P2u - mouse
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 24-Sep-1999
C:Accession: A47556
R:Luftig, K.D.; Shiao, A.K.; Brake, A.J.; Julius, D.
Proc. Natl. Acad. Sci. U.S.A. 90, 5113-5117, 1993
A:Title: Expression cloning of an ATP receptor from mouse neuroblastoma cells.
A:Reference number: A47556; PMID:93281707; PMID:7685114
A:Accession: A47556
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-373 <LUS>
A:Cross-references: GB:L14751; NID:G309457; PIDN:AAA39871.1; PID:G309458
C:Superfamily: ATP receptor P2u
C:Keywords: transmembrane protein

Query Match 16.4%; Score 308; DB 2; Length 373;
Best Local Similarity 27.2%; Pred. No. 1.9e-19;
Matches 96; Conservative 53; Mismatches 164; Indels 40; Gaps 8;

QY 25 SQTVKIFLSRYTALLVFSALGNILALCLTCQSKRNCTGTYLVHLAVSDILFTVLP 84
Db 28 NEDFKVLLPVSYGVVGVGLCLNVALYIFLCLRLKTNASTTTFMFLAVSDSLYASLP 87
QY 85 GRVVCVVLGSSWPFKGLCLRLTAFLVLYTDTYGGVYLMACVSDHYPAV-----CAHWGPR 140
Db 88 LLVYVYARGDHPFSTVLCVLFVLYTNLYCSILFELTCLSVHRCGLVLRPLHSLRW-- 145
QY 141 LRTAGARLVCAVAITWLVLLQTMPLMLMPKFLVGLKACMEYSSMESVGLGLPLMVLVAF 200
Db 146 --EARYARRVAAVVWVVLACQAPVLYFVTTSVRGTRITCHDTSARELSFHFVAYSSVML 203
QY 201 AIGFCGPVGIILSCYMKITWKLCTAREDPVTSRKGHRWGCLLTLMLVAV-VVCFSPY 259
Db 204 GLLFVAVPFSVILVCVYLMARLLKPAY--GTTGGLPRAKRSVTRTALVLAVALCLFLPF 261
QY 260 HLNKQFMARGMLHPLSCAERRAFLLSLQATVALMNMCGITPIIYFFASTHYRWKL--- 316
Db 262 HVTRTLIYFSRDLT--SCHTNAINMAYKITRPLAGANSLCDPLVFLAGQRLVRPADA 320
QY 317 -----LGILK-----LKGSSSSSSSSSSSTPGKASSSTPSI 346
Db 321 KPTEPTPSPQARKGLHRPNRTVRKDLVSDDSRRTSTP--AGSETKDI 371

chemokine (C-C) receptor 1 - human	
N:Alternate names: C-C CKR-1; macrophage inflammatory protein-1-alpha receptor	
C:Species: Homo sapiens (man)	
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 13-Aug-1999	
C:Accession: A45177; I55671	
R:Neote, K.; DiGregorio, D.; Mak, J.Y.; Horuk, R.; Schall, T.J.	
Cell 72, 415-425, 1993	
A>Title: Molecular cloning, functional expression, and signaling characteristics of a C-	
A:Reference number: A45177; MUID:93161416; PMID:7679328	
A:Accession: A45177	
A>Status: nucleic acid sequence not shown	
A:Molecule type: mRNA	
A:Residues: 1-355 <NEO>	
A:Cross-references: GB:L10918; NID:G292416; PIDN:AAA36543.1; PID:G292417	
A:Experimental source: HL60 cells	
A>Note: sequence extracted from NCBI backbone (NCBIP:124876)	
R:Gao, J.	
J. Exp. Med. 177, 1421-1427, 1993	
A>Title: Structure and functional expression of the human macrophage inflammatory 1 alpha	
A:Reference number: I55671; MUID:93240122; PMID:7683036	
A:Accession: I55671	
A>Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: mRNA	
A:Residues: 1-355 <RES>	
A:Cross-references: GB:L10918; NID:G292416; PIDN:AAA36543.1; PID:G292417	
C:Genetics:	
A:Gene: GDB:CMKRI; CMKR-1	
A:Cross-references: GDB:138446; OMIM:601159	
A:Map position: 3p21-3p21	
C:Superfamily: vertebrate rhodopsin	
C:Keywords: disulfide bond; G protein-coupled receptor; glycoprotein; phosphoprotein; tr	
F:36-60/Domain: transmembrane #status predicted <TM>	
F:71-91/Domain: transmembrane #status predicted <TM>	
F:108-129/Domain: transmembrane #status predicted <TM>	
F:147-171/Domain: transmembrane #status predicted <TM>	
F:205-223/Domain: transmembrane #status predicted <TM>	
F:240-264/Domain: transmembrane #status predicted <TM>	
F:288-305/Domain: transmembrane #status predicted <TM>	
F:5/Binding site: carbohydrate (Asn) (covalent) #status predicted	
F:24-273,106-183/Disulfide bonds: #status predicted	
F:345/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicted	
Query Match 16.4%; Score 307; DB 2; Length 355;	
Best Local Similarity 27.3%; Pred. No. 2.2e-19;	
Matches 86; Conservative 58; Mismatches 143; Indels 28; Gaps 5;	
QY	37 YTALLVFSALGNILALCLTCQSRKINCTGIVLVHVSDDLFTVALPGRVVCYVIGSSW 96
Db	37 YTALLVFSALGNILALCLTCQSRKINCTGIVLVHVSDDLFTVALPGRVVCYVIGSSW 96
QY	41 YSLVFIYGLVGNILVYLVLVQYKLNMTSIYLLNLAISDLLFTLP-FWIDYKLDW 99
Db	41 YSLVFIYGLVGNILVYLVLVQYKLNMTSIYLLNLAISDLLFTLP-FWIDYKLDW 99
QY	97 PFQGLCLRTAFVLYTDTYGGVYLMACVSDHYPAVVCVCAHWGPRLTAGRLVCAVWT 156
Db	97 PFQGLCLRTAFVLYTDTYGGVYLMACVSDHYPAVVCVCAHWGPRLTAGRLVCAVWT 156
QY	100 VFDGCKILSGFYITGLYSEIFFILLIDRYLAIVHAFALRARTVTFGVITSLIWA 159
Db	100 VFDGCKILSGFYITGLYSEIFFILLIDRYLAIVHAFALRARTVTFGVITSLIWA 159
QY	157 LVLLQTMPLLMPTKPLVGLKACMEYSMESVGLGLP-LVLAFAIGCGPVGIIISCY 215
Db	157 LVLLQTMPLLMPTKPLVGLKACMEYSMESVGLGLP-LVLAFAIGCGPVGIIISCY 215
QY	160 LAILASMPGLYFSGKTQWEFTHTCSLHFFHESLREWKLPQALKALFGVLVPLLYIICY 219
Db	160 LAILASMPGLYFSGKTQWEFTHTCSLHFFHESLREWKLPQALKALFGVLVPLLYIICY 219
QY	216 MKITWKLCSAREDPVTSRGRHWRGCLLTLLMVAVVCFSPYHNIKQFMARGMLHLP 275
Db	216 MKITWKLCSAREDPVTSRGRHWRGCLLTLLMVAVVCFSPYHNIKQFMARGMLHLP 275
QY	220 TGIKILLRPNE-----KSKAVR---LIFVIMITFFLFWFYNYLTILISVFQDFLTH 271
Db	220 TGIKILLRPNE-----KSKAVR---LIFVIMITFFLFWFYNYLTILISVFQDFLTH 271
QY	276 SCARRAFLLSLQATVALMMNCGITPIIYFFASTHYRWL-----L 317
Db	276 SCARRAFLLSLQATVALMMNCGITPIIYFFASTHYRWL-----L 317
QY	272 ECEQSRHLDLAVQVTEIVATYHCCVNFVIAFYAFGERFRKYLRLQLFHRRVAVHLVKWLPPL 331
Db	272 ECEQSRHLDLAVQVTEIVATYHCCVNFVIAFYAFGERFRKYLRLQLFHRRVAVHLVKWLPPL 331
QY	318 GILKXGSSSSSSSS 332
Db	318 GILKXGSSSSSSSS 332
QY	332 SVDRLEKRVSTSPST 346
Db	332 SVDRLEKRVSTSPST 346
RESULT 14	
A53611	
interleukin-8 receptor type B - human	
C:Species: Homo sapiens (man)	
C>Date: 07-Oct-1994 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999	
C:Accession: I37898; I38712; A53611; A39446	
R:Ahuja, S.K.; Shetty, A.; Tiffany, H.L.; Murphy, P.M.	
J. Biol. Chem. 269, 26381-26389, 1994	
A>Title: Comparison of the genomic organization and promoter function for human interleu	
A:Reference number: I37898; MUID:95014476; PMID:7929358	
A:Accession: I37898	
A>Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-360 <RES>	
A:Cross-references: EMBL:U11869; NID:G511801; PIDN:AAB60656.1; PID:G511803	
A:Accession: I38712	
A>Status: preliminary	
A:Molecule type: mRNA	
A:Residues: 1-15 <RE2>	
A:Cross-references: EMBL:U11872; NID:G511808; PIDN:AAA64380.1; PID:G511809; EMBL:U11873;	
I1876; NID:G511816; PID:G511817; EMBL:U11877; NID:G511818; PID:G511819; EMBL:U11878; NID	
R:Sprenger, H.; Lloyd, A.R.; Lautens, L.L.; Bonner, T.I.; Kelvin, D.J.	
J. Biol. Chem. 269, 11065-11072, 1994	
A>Title: Structure, genomic organization, and expression of the human interleukin-8 rece	
A:Reference number: A53611; MUID:94209273; PMID:7512557	
A:Accession: A53611	
A>Status: preliminary	
A:Molecule type: DNA	
A:Residues: 6-360 <SPR>	
A:Cross-references: GB:M99412; GB:L19593	
R:Murphy, P.M.; Tiffany, H.L.	
Science 253, 1280-1283, 1991	
A>Title: Cloning of complementary DNA encoding a functional human interleukin-8 receptor	
A:Reference number: A39446; MUID:91368200; PMID:1891716	
A:Accession: A39446	
A>Status: preliminary; nucleic acid sequence not shown	
A:Molecule type: mRNA	
A:Residues: 6-360 <MR>	
A:Cross-references: GB:M73969	
C:Comment: This receptor, unlike IL8RA, binds several peptides besides interleukin-8, in	
C:Genetics:	
A:Gene: GDB:IL8RB; IL8RA	
A:Cross-references: GDB:127868; OMIM:146928	
A:Map position: 2q35-2q35	
C:Superfamily: vertebrate rhodopsin	
C:Keywords: G protein-coupled receptor; transmembrane protein	
Query Match 16.3%; Score 306.5; DB 2; Length 360;	
Best Local Similarity 28.0%; Pred. No. 2.5e-19;	
Matches 94; Conservative 61; Mismatches 144; Indels 37; Gaps 9;	
QY	24 ESQTVKIFLSFYTALLVFSALGNILALCLTCQSRKINCTGIVLVHVSDDLFTVAL 83
Db	24 ESQTVKIFLSFYTALLVFSALGNILALCLTCQSRKINCTGIVLVHVSDDLFTVAL 83
QY	42 ESLEINKYFVITIIYALVFLSLGNSLVMLVILYSRGSRVTDVILNLALADLLFALT 101
Db	42 ESLEINKYFVITIIYALVFLSLGNSLVMLVILYSRGSRVTDVILNLALADLLFALT 101
QY	84 PGRVVCYVIGSSWPFQKGLCLRTAFVLYTDTYGGVYLMACVSDHYPAVVCVCAHWGPRLRT 143
Db	84 PGRVVCYVIGSSWPFQKGLCLRTAFVLYTDTYGGVYLMACVSDHYPAVVCVCAHWGPRLRT 143
QY	102 PINAAKNG--WIFGTFCLKVSVLLKEVNFYSGILLACISVDRLAIVHA---TFTLT 156
Db	102 PINAAKNG--WIFGTFCLKVSVLLKEVNFYSGILLACISVDRLAIVHA---TFTLT 156
QY	144 AGR--ARLVCAVATWLTLLQTMPLLLPMT-----KPLVGLKACMEYSMESVGLGLP 195
Db	144 AGR--ARLVCAVATWLTLLQTMPLLLPMT-----KPLVGLKACMEYSMESVGLGLP 195
QY	157 QKRYLVKFCISLWGLSLALPVLFRRTVYSSNVS-----ACEDMGNTANWEMLL 211
Db	157 QKRYLVKFCISLWGLSLALPVLFRRTVYSSNVS-----ACEDMGNTANWEMLL 211
QY	196 VLVAFAIGCGPVGIIISCY---MKITWKLCSAREDPVTSRGRHWRGCLLTLLMLVAV 252
Db	196 VLVAFAIGCGPVGIIISCY---MKITWKLCSAREDPVTSRGRHWRGCLLTLLMLVAV 252
QY	212 RILPQSGFIVPLLIIMLCFYGFTLTLTK-----AHGQKRAMRVFAVVLIF 260
Db	212 RILPQSGFIVPLLIIMLCFYGFTLTLTK-----AHGQKRAMRVFAVVLIF 260
QY	253 VCVSPHYL-NIKQFMARGMLHLPSCARRAFLLSLQATVALMMNCGITPIIYFFASTH 311
Db	253 VCVSPHYL-NIKQFMARGMLHLPSCARRAFLLSLQATVALMMNCGITPIIYFFASTH 311
QY	261 LLCWLYNLVLLADTLNRTQVIQETCERNHIDRALDATEILGILHSLCNPLIYAFIQK 320
Db	261 LLCWLYNLVLLADTLNRTQVIQETCERNHIDRALDATEILGILHSLCNPLIYAFIQK 320
QY	312 YRKWLLGILKLGSSSSSS---SSSTPGKASSET 343
Db	312 YRKWLLGILKLGSSSSSS---SSSTPGKASSET 343
QY	321 FRHGLKILAIHGLIKDLSLPKDSRPSFVGVSSSGHT 356
Db	321 FRHGLKILAIHGLIKDLSLPKDSRPSFVGVSSSGHT 356

RESULT 15
I38435
angiotensin receptor homolog APJ - human
C;Species: Homo sapiens (man)
C;Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 21-Jul-2000
C;Accession: I38435
R;O'Dowd, B.F.; Heiber, M.; Chan, A.; Heng, H.H.; Tsui, L.; Kennedy, J.L.; Shi, X.; Petz
Gene 136, 355-360, 1993
A;Title: A human gene that shows identity with the gene encoding the angiotensin receptor
A;Reference number: I38435; MUID:94124031; PMID:8294032
A;Accession: I38435
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-380 <RES>
A;Cross-references: EMBL:U03642; NID:G425351; PIDN:AAA18954.1; PID:G425352
C;Genetics:
A;Gene: APJ
A;Map position: 11q12
A;Introns: #status absent
C;Superfamily: vertebrate rhodopsin

Query Match	16.1%;	Score 302;	DB 2;	Length 380;
Best Local Similarity	26.7%;	Pred. No. 6.5e-19;		
Matches	88;	Conservative 67;	Mismatches 152;	Indels 22; Gaps 9;

Qy

37

Y T A L L V F S A L G N I L A L C L T C Q K R - K I N C T G I Y L V H L A V S D L L F T V A L P G R V V C Y V L G S S

95

Db

35

Y M L V F L I L G T T G N G I V L W T V F R S S R E K R R S A D I F I A S L A V A D L T F V V T L P L W A T Y T Y R D Y D

94

Qy

96

W P E K G G L C R L T A F V L Y T D T Y G G V Y L M A C V S D H Y P A V C A H W G P R L R T A G R A R L V C V A I W

155

Db

95

W P F G T F C K L S S Y L I F V N M T A S V F C L T G L S F D R Y L A I V R P V A N A R L R L R V S G A V A T A V L W

154

Qy

156

T L V L L Q T M P L L L M P M T K P L - - V G K L A C - M E Y S S M E S V - - - - - L G L P L M V L V A F A I G F C

205

Db

155

V L A A L L A M P V M V L R T T G D L E N T T K V Q C Y M D Y S M V A T V S S E W A W E V G L G V S - - - S T T V G F V

211

Qy

206

G P V G I I L S C Y M K I T W K L C S T A R E D P V T S R K G R H W R G C L L T L L M L V A V - - V V C F S P Y H L N I

263

Db

212

V P F T I M L T C Y F F I A Q T I A G H F R K E R I - - - E G L R K R R L L S I I V L V V T F A L C W M P H L V K

268

Qy

264

K Q F M A R G M L H P S C A E R A F I L S L - Q A T V A L M N N C G I T P I I Y F F A S T H Y R K W L L G I L K L

322

Db

269

T L Y M L G S I L H W P - - C D F D L F I L M N I F P Y C T C I S Y V N S C L N P F L Y A F D P R F R Q A C T S M L C C

326

Qy

323

K G S S S S S S S S S S T P G K A S S E T P S I T Q A R G

351

Db

327

Q Q S R C A G T S H S S G E K G A S Y S S G H S Q G P G

355

Search completed: February 19, 2004, 19:34:56
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 19:18:00 ; Search time 17 Seconds
(without alignments)
995.860 Million cell updates/sec

Title: US-10-041-615-34

Perfect score: 1876
Sequence: 1 MIKLGYPPTPTKTKIKD.....SETPSITQARGSMFLAEHV 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	632.5	33.7	361	EBI2_HUMAN	P32249 homo sapien
2	391.5	20.9	367	GLP17_HUMAN	Q13304 homo sapien
3	367	19.6	428	PAR1_CRLO	Q00991 cricetus
4	358.5	19.1	425	PAR1_FAPHA	P56488 papio hamad
5	357	19.0	430	PAR1_MOUSE	P30558 mus musculus
6	352.5	18.8	425	PAR1_HUMAN	P25116 homo sapien
7	341	18.2	308	V206_FQPV	Q91529 fowlpox vir
8	339	18.1	432	PAR1_RAT	P28824 rattus norv
9	332.5	17.7	420	PAR1_XENLA	P47749 xenopus lae
10	332	17.7	537	P278_XENLA	P79928 xenopus lae
11	329	17.5	328	P276_HUMAN	Q15077 homo sapien
12	326	17.4	353	CKR8_MOUSE	P56484 mus musculus
13	324	17.3	355	CKR8_HUMAN	P51685 homo sapien
14	323.5	17.2	340	CLT1_PIG	Q95N02 sus scrofa
15	322.5	17.2	360	CKR4_HUMAN	P51679 homo sapien
16	322	17.2	337	CLT1_HUMAN	Q92771 homo sapien
17	318	17.0	339	CLT1_RAT	Q92478 rattus norv
18	317	16.9	355	CKR1_MACMU	P56482 macaca mula
19	316.5	16.9	333	FKY9_HUMAN	Q9bxc1 homo sapien
20	316	16.8	352	CLT1_MOUSE	Q991a4 mus musculus
21	316	16.8	373	CM11_HUMAN	Q99788 homo sapien
22	315.5	16.8	371	CM11_MOUSE	P97468 mus musculus
23	308	16.4	328	P276_RAT	Q63371 rattus norv
24	308	16.4	345	CLT2_PIG	Q95N03 sus scrofa
25	308	16.4	373	P272_MOUSE	Q35383 mus musculus
26	307.5	16.4	353	IL8B_PANTR	Q28807 pan troglod
27	307.5	16.4	367	RDG1_HUMAN	P41231 homo sapien
28	307.5	16.4	377	P272_HUMAN	P41231 homo sapien
29	307	16.4	355	CKR1_HUMAN	P32246 homo sapien
30	306.5	16.3	360	IL8B_HUMAN	P25025 homo sapien
31	305.5	16.3	374	P272_RAT	P41232 rattus norv
32	304.5	16.2	353	IL8B_MACMU	Q28519 macaca mula
33	302	16.1	367	CCR3_MOUSE	Q88410 mus musculus

34 302 16.1 380 1 APJ_HUMAN P35414 homo sapien
35 301 16.0 350 1 IL8A_GORGO P55919 gorilla gor
36 300.5 16.0 371 1 CML1_RAT O35786 rattus norv
37 299 15.9 368 1 CCR3_HUMAN P49682 homo sapien
38 298.5 15.9 358 1 IL8B_RABIT P35344 oryctolagus
39 298.5 15.9 359 1 IL8B_MOUSE P35343 mus musculus
40 298 15.9 360 1 IL8B_BOVIN Q28003 bos taurus
41 298 15.9 370 1 P2Y9_HUMAN Q99677 homo sapien
42 297.5 15.9 308 1 P2Y5_CHICK P32250 gallus gall
43 297.5 15.9 353 1 IL8B_GORGO Q28422 gorilla gor
44 297 15.8 355 1 IL8A_RABIT P21109 oryctolagus
45 297 15.8 380 1 APJ_MACMU O97666 macaca mula

ALIGNMENTS

RESULT 1
EBI2_HUMAN STANDARD; PRT; 361 AA.
AC P32249;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE EBV-induced G protein-coupled receptor 2 (EBI2).
GN EBI2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93188173; PubMed=8383238;
RA Birkenbach M.P., Josefsen K., Valamanchili R.R., Lenoir G.M.,
RA Kieff E.;
RT "Epstein-Barr virus-induced genes: first lymphocyte-specific G
RT protein-coupled peptide receptors.";
RL J. Virol. 67:2209-2220(1993).
CC -!- FUNCTION: ORPHAN RECEPTOR. PROBABLE MEDIATOR OF EBV EFFECTS ON B
CC LYMPHOCYTES OR OF NORMAL LYMPHOCYTE FUNCTIONS.
CC -!- TISSUE SPECIFICITY: B-LYMPHOCYTE CELL LINES.
CC -!- INDUCTION: BY EBV.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-5 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L08177; AAA35924.1; -;
CC PIR; B45680; B45680.
CC Genew; HGNC:3128; EBI2.
CC MIM; 605741; -;
CC GO; GO:0005887; C:integral to plasma membrane; TAS.
CC GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; TAS.
CC GO; GO:0006955; P:immune response; TAS.
CC InterPro; IPR000276; GPCR_Rhodpsn.
CC Pfam; PF00001; 7tm.1.1;
CC PRINTS; PR00237; GPCRHOOPS.
CC PROSITE; PS00237; G-PROTEIN_RECEP_F1_1; 1.
CC PROSITE; PS00262; G-PROTEIN_RECEP_F1_2; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein.
CC DOMAIN 1 31 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 32 57 1 (POTENTIAL).
CC DOMAIN 58 77 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 78 95 2 (POTENTIAL).
CC DOMAIN 96 105 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 106 127 3 (POTENTIAL).
CC DOMAIN 128 149 CYTOPLASMIC (POTENTIAL).

```
FT TRANSMEM 150 168 4 (POTENTIAL).
FT DOMAIN 169 192 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 193 215 5 (POTENTIAL).
FT DOMAIN 216 241 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 242 265 6 (POTENTIAL).
FT DOMAIN 266 287 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 288 312 7 (POTENTIAL).
FT DOMAIN 313 361 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 312 361 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 126 134 INTERACTION WITH G PROTEINS.
FT DISULFID 104 181 BY SIMILARITY.
SQ SEQUENCE 361 AA; 41224 MW; B5A217F34C9C67B CRC64;

Query Match 33.7%; Score 632.5; DB 1; Length 361;
Best Local Similarity 37.8%; Pred. No. 5.3e-38;
Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;

QY 2 IKLGFYFTPTTKI-KTKIKDLNVESTVKIFLSRYTALLVFSALGNILALCLTCQKSR 60
Db 3 IQWANNFTPSATPQGNDCDLVAHSTARIVM-PLHYSLVFIIGLVGNLLALVWVQNRK 61
QY 61 KINCTGIYLVHLAVSDLLFTVALPGRVVCVGLSSWPFKGLCRLTAFVLYTDTYGVYL 120
Db 62 KINSTTLYSTNLVISDILFTTALPTRIAYYAGDFDRIGDALCRITALTALFYINTYGVNF 121
QY 121 MACVSDVHPVAVCAHWGPRLTAGRARLVCAVIAWTLVLLQWPLLMPTMTPLVGLKAC 180
Db 122 MTCLSIDRPTAVVHPPLRYNKIKIEHAKGVCFVWLVFAQTLPLLINPMSQEAERITC 181
QY 181 MEYSSMESVUGLPLMVLVPAIGFCOPVGIIILSCYMKITWKLCSAREDPVTSRKGHRW 240
Db 182 MEYFNPEETKSLPWLGLGACFGVLYPLIIILICYSQICCKLFRITAKONPLTEKSGVNNK 241
QY 241 GCLLTLLMVLVAVVCSPPHNLKQFARMG--LHLPSCAERRAFILSLQATVALMMNMC 298
Db 242 ALMTIILIIIVFVLCFTHPVHAIQHWIKLRSNLECSQRHSFOISLHFTVCLMNFVC 301
QY 299 GTPPIYFFASTYRKLWGLILKLGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 352
Db 302 CNDPFIYFFACKYGRKWMELKROVSVISSAVKSAPEENSEMETQTMHSHKSS 358

RESULT 2
GP17 HUMAN STANDARD; PRT; 367 AA.
ID AC Q13304; Q9UDZ6; Q9UE21;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable P2Y purinoceptor GPR17 (P2Y-like receptor) (R12).
GN GPR17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX MEDLINE=96145150; PubMed=8558062;
RA Raport C.J., Schweickart V.L., Chantray D., Eddy R.L. Jr., Shows T.B.,
RA Godiska R., Gray P.W.;
RT "New members of the chemokine receptor gene family.";
RL J. Leukoc. Biol. 59:18-23(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX TISSUE=Hippocampus; PubMed=9523551;
RX MEDLINE=98181695; PubMed=9523551;
RA Blaesus R.H., Weber R.G., Lichter P., Ogilvie A.;
RT "A novel orphan G protein-coupled receptor primarily expressed in the
RT brain is localized on human chromosome band 2q21.";
RL J. Neurochem. 70:1357-1365(1998).
CC -!- FUNCTION: Purative receptor for purines coupled to G-proteins (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
```

```
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q13304-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q13304-2; Sequence=VSP_001987;
CC -!- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U33447; AB16746.1; --
CC EMBL; Y12546; CAAT3144.1; --
CC EMBL; Z94154; CAB08107.1; --
CC EMBL; Z94155; CAB08108.1; --
CC HSSP; P34996; 1DDD.
CC Genew; HGNC:4471; GPR17.
CC MIM; 603071.
CC GO; GO:0005887; C:integral to plasma membrane; TAS.
CC GO; GO:0004950; F:chemokine receptor activity; TAS.
CC GO; GO:0007186; F:G-protein coupled receptor protein signalin. .; TAS.
CC InterPro; IPR000276; GPCR_Rhodopsn.
CC Pfam; PF00001; 7tm 1; 1.
CC PRINTS; PR00237; GPCR_RHODOPSIN.
CC PROSITE; PS00237; G-PROTEIN RECEPT_F1_1; 1.
CC PROSITE; PS0262; G-PROTEIN RECEPT_F1_2; 1.
CC KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Alternative splicing.
KW DOMAIN 1 64 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 65 85 1 (POTENTIAL).
FT DOMAIN 86 92 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 93 113 2 (POTENTIAL).
FT DOMAIN 114 133 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 134 154 3 (POTENTIAL).
FT DOMAIN 155 175 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 176 196 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 197 223 4 (POTENTIAL).
FT TRANSMEM 224 244 5 (POTENTIAL).
FT DOMAIN 245 260 6 (POTENTIAL).
FT TRANSMEM 261 281 7 (POTENTIAL).
FT DOMAIN 282 308 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 309 329 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 330 367 BY SIMILARITY.
FT DISULFID 132 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 42 42 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 204 204 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 282 282 Missing (in isoform 2).
FT VARSPLIC 1 28 /FtId=VSP_001987.
SQ SEQUENCE 367 AA; 40989 MW; 132FBE97BE83C60C CRC64;

Query Match 20.9%; Score 391.5; DB 1; Length 367;
Best Local Similarity 30.5%; Pred. No. 5.8e-21;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;

QY 36 FYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCVGLSS 95
Db 65 FYLLDFILALVGNLTALMLFIRDRHSGTPEANVFLMHLAVADLSCLVLPRLVTHFGNH 124
QY 96 WPFKGLCLRTAFVLYTDTYGVVLMACVSDVHPVAVVCAHWGPRLTAGRARLVCAVIAW 155
Db 125 WPFKGLCLRTAFVLYTDTYGVVLMACVSDVHPVAVVCAHWGPRLTAGRARLVCAVIAW 184
QY 156 TLVLQTVPLLMPTMTPLVGLKACME-YSSNESVLGFLPMVLVPAIGFCOPVGIIILSC 214
Db 185 VVAVAMAPLLVSPQTVQTNHTVWCLQLYREKAS-----HEALVSLAVAFITPITVTC 239
QY 215 YMKITWKLCSAREDPVTSRKGHRWGCLLTLLMVLVAVVCSPPHNLKQFARMGMLHL 274
```

Db 240 YLLATRSRQGRVVE-----RELKTKAVRMIAVLAIPLVCFVYHVNRSYVILHYRSHG 294
 QY 275 PSACBAPFLSLQATVALMNNCGITPIIYFFASTHYRKLGLGIL---KLKG---SSSS 328
 Db 295 ASCATQRIALANRITSCITSLNGALDPIIMFFVAEKERHALCNLCCKRLKGGPPSPFEG 354
 QY 329 SSSSSSTPGKA 339
 Db 355 KTNESLSAKS 365

RESULT 3

PAR1_CRILO STANDARD; PRT; 428 AA.
 AC Q00391; Q60461;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Proteinase activated receptor 1 precursor (PAR-1) (Thrombin receptor).
 GN F2R OR PAR1.
 OS Cricetulus longicaudatus (Long-tailed hamster) (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Cricetulus.
 OX NCBI_TaxID=10030;
 RN [1] SEQUENCE FROM N.A.
 RP TISSUE=Lung;
 RX MEDLINE=91348247; PubMed=1652467;
 RA Rasmussen U.B., Voutret-Craviari V., Jallat S., Schlesinger Y.,
 RA Pages G., Pavirani A., Lecocq J.P., Pouyssegur J.,
 RA Oberghen-Schilling B.;
 RT "cDNA cloning and expression of a hamster alpha-thrombin receptor
 coupled to Ca²⁺ mobilization."
 RL FEBS Lett. 288:123-128(1991).
 RN [2]

SEQUENCE OF 42-428 FROM N.A.
 RX MEDLINE=96028007; PubMed=7488069;
 RA Hartmann T., Grace M.B., Buzard G.S., Russ S.J.;
 RT "Thrombin receptor polymorphism in Chinese hamster."
 RL Biochem. Biophys. Res. Commun. 215:974-980(1995).
 CC -!- FUNCTION: High affinity receptor for activated thrombin coupled to
 CC G proteins that stimulate phosphoinositide hydrolysis.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- PTM: A proteolytic cleavage generates a new amino terminus that
 CC functions as a tethered ligand.
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

DR EMBL; X61958; AAA43957.1; -;
 DR EMBL; U34047; AAA86747.1; -;
 DR HSSP; P34956; 1DDP.
 DR InterPro; IPR000276; GPCR_Rhodopsn.
 DR Pfam; PF00001; 7tm_1; 1.
 DR PRINTS; PR00237; GPCRHOODOPSN.
 DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL1; 1.
 DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 KW Blood coagulation; Polymorphism.
 FT SIGNAL 1 21 POTENTIAL.
 FT PROPEP 22 41 REMOVED FOR RECEPTOR ACTIVATION (BY
 FT SIMILARITY).
 FT CHAIN 42 428 PROTEINASE ACTIVATED RECEPTOR 1.
 FT DOMAIN 42 105 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 106 131 1 (POTENTIAL).

FT DOMAIN 132 140 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 141 160 2 (POTENTIAL).
 FT DOMAIN 161 179 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 180 201 3 (POTENTIAL).
 FT DOMAIN 202 221 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 222 242 4 (POTENTIAL).
 FT DOMAIN 243 271 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 272 291 5 (POTENTIAL).
 FT DOMAIN 292 314 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 315 337 6 (POTENTIAL).
 FT DOMAIN 338 352 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 353 377 7 (POTENTIAL).
 FT DOMAIN 378 428 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 429 448 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 449 468 POLY-SER.
 FT TRANSMEM 469 488 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT SITE 489 508 BY SIMILARITY.
 FT CARBOHYD 509 528 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 529 548 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 549 568 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 569 588 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARIANT 589 608 MISSING (IN AN ALLELE).
 FT CONFLICT 609 628 H -> T (IN REF. 1).
 SQ SEQUENCE 428 AA; 47602 MW; E5492AB2CD877E2F CRC64;
 Query Match 19.6%; Score 367; DB 1; Length 428;
 Best Local Similarity 30.1%; Pred. No. 3.6e-19;
 Matches 102; Conservative 63; Mismatches 158; Indels 16; Gaps 8;
 QY 2 IKLGPYFTPTKIKTKI-KOLN--VESQTVKIFLRSFLTYALLVFSALGNILALCUTCOK 58
 Db 75 IYLNKSHSPAPLAPFISEDAGSLTSPWLRFLFIPSV-YTFVFWLSPLNLAIAVFLK 133
 QY 59 SRKINCTGYLVHLAVSDLLFTVALPGRVVCVVGSSWPFGLGCLTAFVLYTDYGGV 118
 Db 134 MKVKEPAVVYMLHLAVDLVSVLPKISYFSGSDWQFGSGMCFAPFAAFYCNWYASI 193
 QY 119 YLMACVSDHYPAVCAHMGPRIRTAGRARLVCAIWTLLVLTQMPLELLMPMTK--PLVG 176
 Db 194 MLMTVISIDRFVAVVYPIQSLNSWTLGRANFTCLVIMWMAIMGVPLLEKEQTRVPELN 253
 QY 177 KLACMEYSWESVGLPLMVLVAF-AIGCGPVGILSCYMKITWKLCTAREDPVTSRK 235
 Db 254 ITTCHDVNLNETLQGFYSYFAFSAFVFLVLIITSTICYSIIRCLSSSS-----VANR 308
 QY 236 GRHWRCGLTLMLVAVVVCFSYHNLNKFQWARGMLHLPSCAERRAFLLSQATVALMN 295
 Db 309 SKSRALFLSAAFVCFVFCGP--TNVLLIMHYLLSDSPATEKAYFAYLL--CVCVSS 364
 QY 296 MNCGITPIYFFASTHYRKLGLKLGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 334
 Db 365 VSCCIDPLIYVASSCORHLYGILCKESSDPNSYNST 403

RESULT 4

PAR1_PAPHA STANDARD; PRT; 425 AA.
 AC P56488;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Proteinase activated receptor 1 precursor (PAR-1) (Thrombin receptor).
 GN F2R OR PAR1 OR BTHR12.
 OS Papio hamadryas (Hamadryas baboon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Papio.
 OX NCBI_TaxID=9557;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Shoji M., Hayzer D.J., Hanson S.R.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: High affinity receptor for activated thrombin coupled to

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL; M62424; AAA36743.1; --
 EMBL; AF391809; AAK69768.1; --
 EMBL; BC02464; AAK02464.1; --
 PIR; A37912; A37912
 PDB; 1NRN; 31-MAY-94.
 PDB; 1NRQ; 31-MAY-94.
 PDB; 1NRR; 31-MAY-94.
 Genew; HGNC:3537; P2R.
 MIM; 187930; --
 GO; GO:0005794; C:Golgi apparatus; TAS.
 GO; GO:0005887; C:integral to plasma membrane; TAS.
 GO; GO:0005102; F:receptor binding activity; TAS.
 GO; GO:0015057; F:thrombin receptor activity; TAS.
 GO; GO:0006915; P:apoptosis; TAS.
 GO; GO:0007596; P:blood coagulation; TAS.
 GO; GO:0006919; P:caspase activation; TAS.
 GO; GO:0006928; P:cell motility; TAS.
 GO; GO:0007275; P:development; TAS.
 GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
 GO; GO:0007125; P:invasive growth; TAS.
 GO; GO:0000074; P:regulation of cell cycle; TAS.
 GO; GO:0009611; P:response to wounding; TAS.
 GO; GO:0007165; P:signal transduction; TAS.
 GO; GO:0007262; P:STAT protein nuclear translocation; TAS.
 GO; GO:0007260; P:tyrosine phosphorylation of STAT protein; TAS.
 InterPro; IPR000276; GPCR_Rhodopsn.
 Pfam; PF00001; 7tm1.1;
 PRINTS; PR00237; GPCR_Rhodopsn.
 PROSITE; PS00237; G-PROTEIN RECEPTOR FL1; 1.
 PROSITE; PS0262; G-PROTEIN RECEPTOR FL2; 1.
 G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 Blood coagulation; Phosphorylation; Polymorphism; 3D-structure.
 FT SIGNAL 1 26 POTENTIAL.
 FT PROPEP 27 41 REMOVED FOR RECEPTOR ACTIVATION.
 FT CHAIN 42 425 PROTEINASE ACTIVATED RECEPTOR 1.
 FT DOMAIN 42 102 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 103 128 1 (POTENTIAL).
 FT DOMAIN 129 137 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 138 157 2 (POTENTIAL).
 FT DOMAIN 158 176 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 177 198 3 (POTENTIAL).
 FT DOMAIN 199 218 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 219 239 4 (POTENTIAL).
 FT DOMAIN 240 268 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 269 288 5 (POTENTIAL).
 FT DOMAIN 289 311 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 312 334 6 (POTENTIAL).
 FT DOMAIN 335 350 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 351 374 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 375 425 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 426 442 ASP/GLU-RICH (ACIDIC).
 FT SITE 41 42 CLEAVAGE (BY THROMBIN).
 FT DISULFID 175 254 BY SIMILARITY.
 FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 75 75 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 166 166 S -> G (IN dbSNP:5893).
 FT FTID=VAR.014167.
 FT CONFLICT 238 238 L -> V (IN REF. 1).
 FT CONFLICT 335 335 A -> V (IN REF. 3).
 FT CONFLICT 364 364 C -> S (IN REF. 1).
 FT SEQUENCE 425 AA; 47440 MW; 41B742A99EEC96AB CRC64;
 SQ

Query Match 18.8%; Score 352.5; DB 1; Length 425;
 Best Local Similarity 29.7%; Pred. No. 3.8e-18;
 Matches 100; Conservative 57; Mismatches 159; Indels 21; Gaps 7;
 QY 25 SQTWKIFRSLFYTALLVPSALGNILALCLTCKSRKINCTGIYVHLAVSDLLFTVALP 84
 DB 98 SSMITLFVPSV-YTGVEFVSLPLNIMAVFVILMKVKGPAVVMGLATADLVFVSLP 156
 QY 85 GRVVCYVLGSSWPFKGLCLRTAFVLYTDYGGVYLMACVSVDPHYPAVCAHWGPRLRTA 144
 DB 157 FKISYFSGSDWQFSGELCFRTAAFCNNMYASILLMTVISIDRFLAVVYPMQSLSWRTL 216
 QY 145 GRARLVCAITWTLVLOTMPLLLMPMTKPLVGLKACMEYSMESVGLPLM-----VLV 198
 DB 217 GRASFTCLAIWALAIAGVWFLLEKQIQVPG----LNITTCHEVLTNETLLEGGYAYYFS 272
 QY 199 AF-AIGFCGPGVGIILSCYMKITKWLKSTAREDPVTSRKGHRGCLLTLLMLVAVVVCFS 257
 DB 273 AFSAVFEFVLLISTVCYVSIIRCLSSA-----VANRSKSRALFLSAVFCIFICFG 327
 QY 258 PYHNIKQFWRGMLHLPSCAERRAFLLSLQATVALNMNMCGITPIIYFFASTHYRKWL 317
 DB 328 P--TNVLLIAHYSFLSHTSTTEAAYFAYLL--CVCVSSISCCIDPLIYYVASSECQRYV 383
 QY 318 GILKLGSSSSSSSSSTPGKASSETPSITQARGSMF 354
 DB 384 SILCKESSDPSSYSSGQLMASHKMDTCSSNLNNSIY 420
 RESULT 7
 V206 FOWPV STANDARD; PRT; 308 AA.
 AC Q9J529;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE G-protein coupled receptor homolog FPV206.
 OS Fowlpox virus (FPV).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Avipoxvirus
 OC NCBI_TaxID=10261;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20193820; PubMed=10729156;
 RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
 RT "The genome of fowlpox virus."
 RL J. Virol. 74:3815-3831(2000).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF198100; AAP44550.1; --
 DR InterPro; IPR000276; GPCR_Rhodopsn.
 DR Pfam; PF00001; 7tm1.1;
 DR PRINTS; PR00237; GPCR_Rhodopsn.
 DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL1; 1.
 DR PROSITE; PS0262; G-PROTEIN RECEPTOR FL2; 1.
 DR G-protein coupled receptor; Transmembrane; Glycoprotein.
 KW DOMAIN 1 20 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 21 41 1 (POTENTIAL).
 FT DOMAIN 42 52 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 53 73 2 (POTENTIAL).
 FT DOMAIN 74 91 2 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 92 112 3 (POTENTIAL).
 FT

```
FT DOMAIN 113 133 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 134 154 4 (POTENTIAL).
FT DOMAIN 155 180 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 181 201 5 (POTENTIAL).
FT DOMAIN 202 226 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 227 247 6 (POTENTIAL).
FT DOMAIN 248 270 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 271 291 7 (POTENTIAL).
FT DOMAIN 292 308 CYTOPLASMIC (POTENTIAL).
FT DISULFID 89 167 BY SIMILARITY.
FT CARBOHYD 2 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 308 AA; 35375 MW; 5ABF53BE1FEB28A CRC64;

Query Match 18.2%; Score 341; DB 1; Length 308;
Best Local Similarity 29.0%; Pred. No. 1.8e-17;
Matches 90; Conservative 55; Mismatches 129; Indels 28; Gaps 8;

QY 34 SLFPTALLVPSA--IGNILALCL-----TCQSRKINCTGIVLVHVAVDLLFTVALPGRV 87
DB 19 SLAVVTILFFSSLLINISALVIGFYTPAGPMK-----MYLINLVSDILFTVTLPLKI 73
QY 88 VCVLGSWPPKGLCRITAFVLYTDYGVVLMACVSDHYPAVCAHWGPRLRTAGRA 147
DB 74 DYVYFFNWRGEMACRIMSFSLVINTVSNFMFTWISVNYAYVTRPHKYNSTRDNIWRT 133
QY 148 RLVCAVATVLLQTPMLLMPT-KPLVGKACMEYSMESVGLPLMLVAFAP-IGFC 205
DB 134 KIACACTVILVPMSSILFVSTSSDHEKIRCEYKNVGDVSMYLPWPVWIVMCFIGFV 193
QY 206 GPVGIILSYMKITWKLSTAREDPVTSRGRHWGECLLTLLVAVVVCSPVHLNIQ 265
DB 194 IFPMAWISYSAVCTVLSG-----ISKSTRSYRTCKLVACILTEFVCLFPHASVIS 247
QY 266 FMARGMLHLP-----CAERAFLLSLQATVALMMNCGITPIIYFFASTHYRWKLGL 320
DB 248 Y-----MIHITSKTVLCENVSYQYMLLHATQCLMKLNCMDPIIYLFVSSYKSAKNSI 303
QY 321 KL 322
DB 304 KL 305

RESULT 8
PARI RAT STANDARD; PRT; 432 AA.
AC F26824;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Proteinase activated receptor 1 precursor (PAR-1) (Thrombin receptor).
GN F2R OR PARI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Rortic smooth muscle;
RA MEDLINE=92281002; PubMed=1324917;
RX Zhong C., Hayzer D.J., Corson M.A., Wick K., Runge M.S.;
RT "Molecular cloning of the rat vascular smooth muscle thrombin
RT receptor. Evidence for in vitro regulation by basic fibroblast growth
RT factor."
RL J. Biol. Chem. 267:16975-16979(1992).
CC -!- FUNCTION: High affinity receptor for activated thrombin coupled to
CC G proteins that stimulate phosphoinositide hydrolysis.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- PTM: A proteolytic cleavage generates a new amino terminus that
CC functions as a tethered ligand.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M81642; AAA42274.1; -.
DR PIR; A43448; A43448.
DR HSP; P34996; 1DDD.
DR InterPro; IPR000276; GPCR_Rhodpen.
DR Pfam; PF00001; 7tm 1; 1.
DR PRINTS; PR00237; GPCR_Rhodopsin.
DR PROSITE; PS00237; G PROTEIN RECP Fl.1; 1.
DR PROSITE; PS00262; G PROTEIN RECP Fl.2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Blood coagulation.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 45 REMOVED FOR RECEPTOR ACTIVATION (BY
FT CHAIN 46 432 SIMILARITY).
FT DOMAIN 46 109 PROTEINASE ACTIVATED RECEPTOR 1.
FT TRANSMEM 110 135 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 136 144 1 (POTENTIAL).
FT TRANSMEM 145 164 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 165 183 2 (POTENTIAL).
FT TRANSMEM 184 205 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 206 225 3 (POTENTIAL).
FT TRANSMEM 226 246 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 247 275 4 (POTENTIAL).
FT TRANSMEM 276 295 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 296 318 5 (POTENTIAL).
FT TRANSMEM 319 341 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 342 357 6 (POTENTIAL).
FT TRANSMEM 358 381 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 382 432 7 (POTENTIAL).
FT DOMAIN 64 68 CYTOPLASMIC (POTENTIAL).
FT SITE 87 93 ASP/GLU-RICH (ACIDIC).
FT SITE 45 46 POLY-PRO.
FT DISULFID 182 261 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT CARBOHYD 69 69 BY SIMILARITY.
FT CARBOHYD 82 82 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 432 AA; 48280 MW; DD032B97ABA4A605 CRC64;

Query Match 18.1%; Score 339; DB 1; Length 432;
Best Local Similarity 29.0%; Pred. No. 3.5e-17;
Matches 98; Conservative 61; Mismatches 153; Indels 26; Gaps 9;

QY 6 PYFTPTKTKTKIDLN--VESQTVKIFLRSFYTALLVFSALGNIALCLCTQSKRKIN 63
DB 87 PMPPPPPTIS--EDASGYLTSPWLTLPFSV-YTFVFIIVSLPLNLAIAVFFVRMKVK 142
QY 64 CTGTYLVHVAVDLLFTVALPGRVVCVYLGSSWPPFGKLCRLTAFVLYTDTTGGVYLMAC 123
DB 143 PAVVYMLHLMADVLVSVLPFKISYIFSGTDQWFGSGMCFRATACYNMYASIMLTV 202
QY 124 VSDHYPAVCAHWGPRLRTAGRAVLCVAIWTLLQTMPLLMPTKPLVGKLACMEY 183
DB 203 ISIDRFVAVVYPIQSLSWTLGRANTCVIWMALMGVVPVLLKQTTQVPG---LNI 258
QY 184 SSMESVGLPLM-----VLVAF-AIGFCGPVGIILSCYMKITWKLCTSTAREDPVTSRKG 236
DB 259 TTCHDVLNETLLHGFYSYVFSAFSAIFFLVPLTIISTVCTYSITRCLSSSA-----VANRS 313
QY 237 RHWGCLLTLLMVAVVVCFSPVHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMMN 296
DB 314 KGRALFLSAANVCFIIVCGFPTNVLL---IVHLLSDSPGTETAFYAILLCVCVTVA 370
QY 297 NCGITPIIYFFASTHYRWKLGLKLGSSSSSSSSSS 334
DB 371 SC-IDPLIVYVASSECKHLYSILCKRESSDSNSCNS 407
```

```
Query Match 17.7%; Score 332.5; DB 1; Length 420;
Best Local Similarity 26.6%; Pred. No. 9.7e-17;
Matches 92; Conservative 74; Mismatches 155; Indels 25; Gaps 8;

QY 16 TKIKDLNVEQTKVIFLRSFYFALLVFSALGNILALCLTCQSKRINKTGIYLVHLAVS 75
DB 88 TKEAQYLSQWLTKEFVSL-YTVWFIVGLPLNLLAIIFLFQKVRKPAVVMNLAI 146
QY 76 DLLFTVALPRVYVYVGLSSWPFKGLCLRTALVLTDTYGVYLMACVSDHYPAVCA 135
DB 147 DVFFSVLPFKIAHYHSGNDWLFPGMCRIVTAIFVCMYCVLLIASISVDRELA 206
QY 136 HWGPRITAGRLVCVAIWTLVLLQTMPLLMPTKPLVGLKACMEYMSVGLPLM 195
DB 207 MHSLSWRTMSRAYMACSFIWLSIASITPLLVTEQIQ---KIPRLDITCHVDL 262
QY 196 --VIVAFAGFC---GPGVILSCVMKITKLCSTAREDPVTSRKGHRWGCCLTLL 248
DB 263 KDFIYFSSFCLLFFVFPFIITICIGIIRLSSSSIENSC-----KKTALF 317
QY 249 LVAVVVCFSYHINIKQFMARGMLHLPSCARRAFLLSLQATVALMNMCGITPI 308
DB 318 LCVFIICFGTNN---LFLTH---VLQEAERFLYFAYILSACVG--SVSCCL 369
QY 309 STHYKWLILGLKLGSSSSSSSSSSSTPGKASSETPSITQARGSMF 354
DB 370 SSCQRYLSLCCRKVSEPGSGTGMSTAMKNDNCSTNAKSSII 415

RESULT 10
P2Y8 XENLA STANDARD; PRT; 537 AA.
ID "P2Y8 XENLA STANDARD; PRT; 537 AA.
AC P7928;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE P2Y purinoreceptor 8 (P2Y8).
OS Xenopus.
GN Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
NCBI_TaxID=8355;
RX MEDLINE=97284734; PubMed=9139711;
RA Bogdanov Y.D., Dale L., King B.F., Whittock N., Burnstock G.;
RT "Early expression of a novel nucleotide receptor in the neural plate
of Xenopus embryos.";
RL J. Biol. Chem. 272:12583-12590(1997).
CC 1- FUNCTION: RECEPTOR FOR EXTRACELLULAR ATP, UTP, CTP, GTP AND ITP.
CC THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH
CC ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
CC MAY PLAY A KEY ROLE IN THE EARLY DEVELOPMENT OF NEURAL TISSUE.
CC 1- SUBCELLULAR LOCATION: Integral membrane protein.
CC 1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; X99953; CAA68213.1; --
DR HSP; P34996; 1DD.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODOPS.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL1; 1.
DR PROSITE; PS0262; G-PROTEIN RECEPTOR FL2; 1.
DR PROSITE; PS50262; G-PROTEIN RECEPTOR FL2; 1.

PARI_XENLA STANDARD; PRT; 420 AA.
ID PARI_XENLA STANDARD; PRT; 420 AA.
AC P47749;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Proteinase activated receptor 1 precursor (PAR-1) (Thrombin receptor).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
NCBI_TaxID=8355;
RX MEDLINE=94195429; PubMed=8145852;
RA Gersten R.E., Chen J., Ishii M., Ishii K., Nanavicz T.,
RA Turk C.W., Vu T.-K.H., Coughlin S.R.;
RT "Specificity of the thrombin receptor for agonist peptide is defined
by its extracellular surface.";
RL Nature 368:648-651(1994).
CC 1- FUNCTION: High affinity receptor for activated thrombin coupled to
CC G proteins that stimulate phosphoinositide hydrolysis.
CC 1- SUBCELLULAR LOCATION: Integral membrane protein.
CC 1- PTM: A proteolytic cleavage generates a new amino terminus that
CC functions as a tethered ligand.
CC 1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; U09632; AAA18498.1; --
DR PIR; I51667; 1S1667.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODOPS.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL1; 1.
DR PROSITE; PS0262; G-PROTEIN RECEPTOR FL2; 1.
DR PROSITE; PS50262; G-PROTEIN RECEPTOR FL2; 1.
KW Blood coagulation.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 42 REMOVED FOR RECEPTOR ACTIVATION.
FT CHAIN 43 420 PROTEINASE ACTIVATED RECEPTOR 1.
FT DOMAIN 43 101 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 102 127 1 (POTENTIAL).
FT DOMAIN 128 136 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 137 156 2 (POTENTIAL).
FT DOMAIN 157 175 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 176 197 3 (POTENTIAL).
FT DOMAIN 198 217 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 218 238 4 (POTENTIAL).
FT DOMAIN 239 267 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 268 287 5 (POTENTIAL).
FT DOMAIN 288 310 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 311 333 6 (POTENTIAL).
FT DOMAIN 334 345 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 346 369 7 (POTENTIAL).
FT DOMAIN 370 420 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 295 300 POLY-SER.
FT SITE 42 43 CLEAVAGE (BY THROMBIN).
FT DISULFID 174 253 BY SIMILARITY.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT MUTAGEN 42 42 K-A: LACK OF ACTIVATION BY THROMBIN.
SQ SEQUENCE 420 AA; 47435 MW; D5163F56AFE12372 CRC64;
```

MEDLINE=97432828; PubMed=9286708;
Somers G.R., Hammet F., Woollett E., Richards R.I., Southey M.C.,
Venter D.J.;
"Chromosomal localization of the human P2y6 purinoceptor gene and
phylogenetic analysis of the P2y purinoceptor family";
Genomics 44:127-130(1997).
[3]
SEQUENCE FROM N.A.
TISSUE=Placenta;
MEDLINE=97415792; PubMed=9268704;
Maler R., Glatz A., Moshbacher J., Bilbe G.;
"Cloning of P2y6 cDNAs and identification of a pseudogene: comparison
of P2Y receptor subtype expression in bone and brain tissues";
Biochem. Biophys. Res. Commun. 237:297-302(1997).
[4]
TISSUE=Brain;
MEDLINE=98069816; PubMed=9412455;
Maler R., Glatz A., Moshbacher J., Bilbe G.;
Biochem. Biophys. Res. Commun. 240:298-302(1997).
[5]
SEQUENCE FROM N.A.
Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
Yu J., Han L.H.;
"Novel human cDNA clones with function of inhibiting cancer cell
growth";
Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
[6]
SEQUENCE FROM N.A.
Puhl H.L. III, Ikeda S.R., Aronstam R.S.;
Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
[7]
SEQUENCE FROM N.A.
TISSUE=Brain;
MEDLINE=423386257; PubMed=12477932;
Straussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shermen C.M., Schuler G.D.,
Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Datchenko L., Marudina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Frange C.,
Rana S.S., McQuellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
Boak S.A., Morway P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Souffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.J., Skalska U., Smalusz D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
-!- FUNCTION: RECEPTOR FOR EXTRACELLULAR UDP > UTP > ATP. THE ACTIVITY
OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE A
PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
-!- SUBCELLULAR LOCATION: Integral membrane protein.
-!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
-!- This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
EMBL; X97058; CAA65770.1; -
EMBL; U52464; AAB03572.1; -
EMBL; AF007892; AAB80713.1; -
EMBL; AF007891; AAB80712.1; -

CC88 OR CMK88 OR TER1.
Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN
RP
SEQUENCE FROM N.A.
RA Zaballos A., Goya I.I., Gutierrez J., Varona R., Marquez G.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN
RP
SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=98334001; PubMed=9670926;
RA Zingoni A., Soto H., Hedrick J.A., Stoppacciaro A., Storlazzi C.T.,
RA Sinigaglia F., D'Amrosio D., O'Garra A., Robinson D., Rocchi M.,
RA Santoni A., Zlotnick A., Napolitano M.,
RA "The chemokine receptor CCR8 is preferentially expressed in Th2 but
RT not Th1 cells.";
RL J. Immunol. 161:547-551(1998).
CC -1- FUNCTION: RECEPTOR FOR THE TCA-3 CHEMOKINE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

CC EMBL; Z98206; CAB10896.1; -;
CC EMBL; Z98205; CAB10895.1; -;
CC EMBL; AF001277; AAC97598.1; -;
CC MGD; MGI:1201402; Ccr8.
CC GO; GO:0016493; F-C-C chemokine receptor activity; IDA.
CC GO; GO:0005515; F-protein binding activity; IPI.
CC GO; GO:0006935; P-chemotaxis; IDA.
CC InterPro; IPR004068; CC_chemokine8.
CC InterPro; IPR000276; GPCR_Rhodpsn.
CC Pfam; PF00001; 7tm 1; 1.
CC PRINTS; PR01530; CHEMOKINER8.
CC PRINTS; PR00237; GPCR_RHODPSN.
CC PROSITE; PS00237; G-PROTEIN RECEPTOR_F1_1; 1.
CC PROSITE; PS00262; G-PROTEIN RECEPTOR_F2_1; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 33 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 34 61 1 (POTENTIAL).
FT DOMAIN 62 71 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 72 91 2 (POTENTIAL).
FT DOMAIN 92 105 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 106 127 3 (POTENTIAL).
FT DOMAIN 128 144 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 145 169 4 (POTENTIAL).
FT DOMAIN 170 200 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 201 220 5 (POTENTIAL).
FT DOMAIN 221 236 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 237 261 6 (POTENTIAL).
FT DOMAIN 262 278 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 279 302 7 (POTENTIAL).
FT DOMAIN 303 353 CYTOPLASMIC (POTENTIAL).
FT DISULFID 104 181 BY SIMILARITY.
FT CARBOHYD 8 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 353 AA; 40045 MW; 31EC4B642CDB9A5 CRC64;
Query Match 17.4%; Score 326; DB 1; Length 353;
Best Local Similarity 25.0%; Pred.No. 2.4e-16;
Matches 91; Conservative 71; Mismatches 134; Indels 68; Gaps 10;
QY 7 YETPTTKIKIDLVNVEQTVKIFLRSFYTALLVFSALGN--ILALCLTCKSRKINC 64
DB 18 FTAPCDAEFLRG-----SMLYALIVYCVFLGLGNSILVILVGVCKLRSI-- 67

CC88 OR CMK88 OR TER1.
Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN
RP
SEQUENCE FROM N.A.
RA Zaballos A., Goya I.I., Gutierrez J., Varona R., Marquez G.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN
RP
SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=98334001; PubMed=9670926;
RA Zingoni A., Soto H., Hedrick J.A., Stoppacciaro A., Storlazzi C.T.,
RA Sinigaglia F., D'Amrosio D., O'Garra A., Robinson D., Rocchi M.,
RA Santoni A., Zlotnick A., Napolitano M.,
RA "The chemokine receptor CCR8 is preferentially expressed in Th2 but
RT not Th1 cells.";
RL J. Immunol. 161:547-551(1998).
CC -1- FUNCTION: RECEPTOR FOR THE TCA-3 CHEMOKINE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; Z98206; CAB10896.1; -;
EMBL; Z98205; CAB10895.1; -;
EMBL; AF001277; AAC97598.1; -;
MD; MGI:1201402; Ccr8.
GO; GO:0016493; F-C-C chemokine receptor activity; IDA.
GO; GO:0005515; F-protein binding activity; IPI.
GO; GO:0006935; P-chemotaxis; IDA.
DR InterPro; IPR004068; CC_chemokine8.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm 1; 1.
DR PRINTS; PR01530; CHEMOKINER8.
DR PRINTS; PR00237; GPCR_RHODPSN.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR_F1_1; 1.
DR PROSITE; PS0262; G-PROTEIN RECEPTOR_F2_1; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 33
FT TRANSMEM 34 61
FT DOMAIN 62 71
FT TRANSMEM 72 91
FT DOMAIN 92 105
FT TRANSMEM 106 127
FT DOMAIN 128 144
FT TRANSMEM 145 169
FT DOMAIN 170 200
FT TRANSMEM 201 220
FT DOMAIN 221 236
FT TRANSMEM 237 261
FT DOMAIN 262 278
FT TRANSMEM 279 302
FT DOMAIN 303 353
FT DISULFID 104 181
FT CARBOHYD 8
FT N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 353 AA; 40045 MW; 31EC4B642CDB9A5 CRC64;
Query Match 17.4%; Score 326; DB 1; Length 353;
Best Local Similarity 25.0%; Pred.No. 2.4e-16;
Matches 91; Conservative 71; Mismatches 134; Indels 68; Gaps 10;
QY 7 YETPTTKIKIDLVNVESTVKVIFLRSFYTALLVFSALGN--ILALCLTCKSRKINC 64
DB 18 FTAPCDAEFLRG-----SMLYALIVYCVFLGLGNSILVILVGVCKLRSI-- 67

QY 26 QTVKIFRLSFLYFALLVFSALGN--ILALCLTCQSKRKINCTGIYLVHLAVSDLLFTVAL 83
 Db 31 QINGKLLAVFYCLLFVSLGNSILVILVWCKLRSL--TDVYLLNLLALSDLLVFSF 88
 QY 84 PGRVVCYVLGSSPCKGKLCRTAFVLYTDTYGVVLMACVSDHYPAVCAHWGPRLRT 143
 Db 89 PFT--YYLDDQWFTVWCKVSGFYIYGFSSFFITLMSVDRLAVHRAVYALKVRT 146
 QY 144 AGRARLVCAIWTLLVLTQMPLLMPTKPLVGLKLAEMEYSMESVGLPLMLVAVAFIG 203
 Db 147 IRMGTTCLLAVLMTATMPTLLVYQVASEDGVLCYSFYNQQT--LKWKFITNFKON 203
 QY 204 FCG---PVGILLSCYMKITWKL--CSTAREDPVTSRKGHRWGCLLTLLMLVAVVVCSP 258
 Db 204 ILGLLPFTTFMFCYIKILHQLKRCQNHKT-----KAIRLVLLIVFASLLFWP 253
 QY 259 YHLNIQKFMARGMLHLPSCAERRAFLLSQATVALMMNMGITPIIYFFASTHYRWL-- 316
 Db 254 FNVVFLTSLHSHMLDGCISQLVYATHVTEIISFTHCVCNPNVIYAPVGEKFKHLSE 313
 QY 317 -----LGIILKLGSSSSSSSSSTPGKASS 341
 Db 314 IFQKSCQIFNYLGRQWPRESCKEKSSCOOHSRSS 350

RESULT 14
 CLT1_PIG STANDARD; PRT; 340 AA.
 AC Q95N02;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cysteinyll leukotriene receptor 1 (CysLTR1).
 GN CysLTR1 OR CysLTR1.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Takasaki J., Kamohara M., Saito T., Matsumoto M., Matsumoto S.,
 RA Onishi T., Soga T., Matsushima H., Furuchi K.,
 RT "Characterization of cloned rat and porcine cysteinyl leukotriene
 RT receptors."
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Receptor for cysteinyl leukotrienes mediating
 CC constriction of the microvascular smooth muscle during an
 CC inflammatory response. This response is mediated via a G-protein
 CC that activates a phosphatidylinositol-calcium second messenger
 CC system (By similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; AB052686; BAB60826.1;
 DR InterPro; IPR004071; Cysleuk receptor.
 DR InterPro; IPR000276; GPCR_Rhodopsn.
 DR Pfam; PF00001; 7tm1_1.
 DR PRINTS; PR01533; CysLTR1.
 DR PRINTS; PR00237; GPCR_Rhodopsn.
 DR PROSITE; PS00237; G-PROTEIN RECEPTOR F1 1; FALSE_NEG.
 DR PROSITE; PS0262; G-PROTEIN RECEPTOR F1 2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.
 FT DOMAIN 1 31
 FT TRANSMEM 32 52
 FT TRANSMEM 53 60
 FT DOMAIN 53 60
 FT CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 61 81
 FT DOMAIN 82 109
 FT TRANSMEM 110 130
 FT DOMAIN 131 144
 FT TRANSMEM 145 165
 FT DOMAIN 166 196
 FT TRANSMEM 197 217
 FT DOMAIN 218 233
 FT TRANSMEM 234 254
 FT DOMAIN 255 279
 FT TRANSMEM 280 300
 FT DOMAIN 301 316
 FT DISULFID 99 176
 FT CARBOHYD 6 6
 FT CARBOHYD 18 18
 FT CARBOHYD 172 172
 FT CARBOHYD 265 265
 SQ SEQUENCE 340 AA; 38986 MW; 54F9372A12CE413 CRC64;

Query Match 17.2%; Score 323.5; DB 1; Length 340;
 Best Local Similarity 27.5%; Pred. No. 3.5e-16;
 Matches 90; Conservative 62; Mismatches 156; Indels 19; Gaps 6;

QY 22 NVESQTVKIF---LRSFLYFALLVFSALGNILALCLTCQSKRKINCTGIYLVHLAVSDLL 78
 Db 15 NTCNDTIDDFRNQVYSTLYSMITVWGFNGFVLYVLIKTYHEKSAYQVVTMINLAVADLL 74
 QY 79 FTVALPGRVVCYVLGSSWPFGKGLCRTAFVLYTDTYGVVLMACVSDHYPAVCAHWG 138
 Db 75 CVCTLPFRVYVYVHKGLWFGDFLCRLSTYALVYVLYCSIFFMTAMSFRCIALVFPVQN 134
 QY 139 PRLTAGARLVCAIWTLLVLTQMPLLMPTKPLVGLKLAEMEYSMESVGLPLMLV 197
 Db 135 INLIYTHKKAKIVCIAIWFILVLTSSPFLMSTYKDKNNTKCFEPPOXNOAKYHVLVLYH 194
 QY 198 VAPAFGCGPVGILLSCYMKITWKLCTAREDPVTSRKGHRWGCLLTLLMLVAVVVCFS 257
 Db 195 VSLFVGHIIIPVLIIVCYTMIITLLKNSMKNLISRK-----KAIGNLIIVTAAFLISEM 250
 QY 258 PYHLNIK---QFMARGMLHLPSCAERRAFLLSQATVALMMNMGITPIIYFFASTHYRK 314
 Db 251 PYHQRTHLHFLHNDTKHCDVLRMQK---SVATILSLAASNCDFDPLLYFFSGGNFRE 307
 QY 315 WLLGILKLGSSSSSSSSSTPGKASS 341
 Db 308 GLSTFRK-----HSLSTMVYVPKKITS 329

RESULT 15
 CKR4_HUMAN STANDARD; PRT; 360 AA.
 AC P51679; Q9ULY6; Q9ULY7,
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE C-C chemokine receptor type 4 (C-CR-4) (CC-CR-4) (CCR-4) (CCR4) (K5-5).
 GN CCR4 OR CCR4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Spleen;
 RX MEDLINE=95370289; PubMed=7642634;
 RA Power C.A., Meyer A., Nemeth K., Bacon K.B., Hoogwerf A.J.,
 RA Proudfoot A.E.I., Wells T.N.C.;
 RT "Molecular cloning and functional expression of a novel CC chemokine
 RT receptor cDNA from a human basophilic cell line."
 RL J. Biol. Chem. 270:19495-19500(1995).
 RN [2]
 RP SEQUENCE FROM N.A., AND VARIANTS VAL-130 AND SER-178.

RA MEDLINE=21040311; PubMed=11196669;
RA Kato H., Tsuchiya N., Izumi S., Miyamasu M., Nakajima T., Kawasaki H.,
RA Hirai K., Tokunaga K.,
RT "New variations of human CC-chemokine receptors CCR3 and CCR4.";
RL Genes Immun. 1:97-104(1999).
RN [3]
RN FUNCTION.
RX MEDLINE=97313486; PubMed=9169480;
RX Imai T., Baba M., Nishimura M., Kakizaki M., Takagi S., Yoshie O.,
RA "The T cell-directed CC chemokine TARC is a highly specific
RT biological ligand for CC chemokine receptor 4.";
RL J. Biol. Chem. 272:15036-15042(1997).
RN [4]
RN FUNCTION.
RX MEDLINE=98104168; PubMed=9430724;
RX Imai T., Chantry D., Raport C.J., Wood C.L., Nishimura M., Godiska R.,
RA Yoshie O., Gray P.W.,
RT "Macrophage-derived chemokine is a functional ligand for the CC
RT chemokine receptor 4.";
RL J. Biol. Chem. 273:1764-1768(1998).
RN [5]
RN FUNCTION.
RX MEDLINE=99394604; PubMed=10466728;
RX Campbell J.J., Haraldsen G., Pan J., Rottman J., Qin S., Ponath P.,
RA Andrew D.P., Wernke R., Ruffing N., Kassam N., Wu L., Butcher E.C.,
RT "The chemokine receptor CCR4 in vascular recognition by cutaneous but
RT not intestinal memory T cells.";
RL Nature 400:776-780(1999).
RN [6]
RN FUNCTION, TISSUE SPECIFICITY, AND PHOSPHORYLATION.
RX MEDLINE=20219238; PubMed=10754297;
RX Ingberdingen M., Dana J.B., Maghazachi A.A.,
RA "Human NK cells express CC chemokine receptors 4 and 8 and respond to
RT thymus and activation-regulated chemokine, macrophage-derived
RT chemokine, and I-309.";
RL J. Immunol. 164:4048-4054(2000).
CC -I- FUNCTION: HIGH AFFINITY RECEPTOR FOR THE C-C TYPE CHEMOKINES
CC TARC/SCYAL7 AND MDC/SCY22. THE ACTIVITY OF THIS RECEPTOR IS
CC MEDIATED BY G(I) PROTEINS WHICH ACTIVATE A PHOSPHATIDYLINOSITOL-
CC CALCIUM SECOND MESSENGER SYSTEM. CAN FUNCTION AS A CHEMOATTRACTANT
CC HOMING RECEPTOR ON CIRCULATING MEMORY LYMPHOCYTES AND AS A
CC CORRECTOR FOR SOME PRIMARY HIV-2 ISOLATES. IN THE CNS, COULD
CC MEDIATE HIPPOCAMPAL-NEURON SURVIVAL.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein.
CC -I- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE THYMUS, IN
CC PERIPHERAL BLOOD LEUKOCYTES, INCLUDING T CELLS, MOSTLY CD4+ CELLS,
CC AND BASOPHILS, AND IN PLATELETS; AT LOWER LEVELS, IN THE SPLEEN
CC AND IN MONOCYTES. DETECTED ALSO IN MACROPHAGES, IL-2-ACTIVATED
CC NATURAL KILLER CELLS AND SKIN-HOMING MEMORY T CELLS, MOSTLY THE
CC ONES EXPRESSING THE CUTANEOUS LYMPHOCYTE ANTIGEN (CLA). EXPRESSED
CC IN BRAIN MICROVASCULAR AND CORONARY ARTERY ENDOTHELIAL CELLS.
CC -I- PTM: IN NATURAL KILLER CELLS, SCY22 BINDING INDUCES
CC PHOSPHORYLATION ON YET UNDEFINED SER/THR RESIDUES, MOST PROBABLY
CC BY BETA-ADRENERGIC RECEPTOR KINASES 1 AND 2.
CC -I- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X85740; CAA59743.1; -
DR EMBL; AB023888; BAA86965.1; -
DR EMBL; AB023889; BAA86966.1; -
DR EMBL; AB023890; BAA86967.1; -
DR EMBL; AB023891; BAA86968.1; -
DR EMBL; AB023892; BAA86969.1; -
DR PIR; A57160; A57160.
DR Genew; HGNC:1605; CCR4.
DR MIM; 604836; -

DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0004950; F:chemokine receptor activity; TAS.
DR GO; GO:0006335; P:chemotaxis; TAS.
DR GO; GO:0007304; P:cytosolic calcium ion concentration elevation; TAS.
DR GO; GO:0006955; P:immune response; TAS.
DR GO; GO:0006954; P:inflammatory response; TAS.
DR InterPro; IPR000276; GPCR_Rhodopsin.
DR Pfam; PF00001; 7tm 1; 1.
DR PRINTS; PR00237; GPCRHHODOPS.
DR PROSITE; PS00237; G-PROTEIN_RECEP_FL_1; 1.
DR PROSITE; PS00262; G-PROTEIN_RECEP_FL_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Phosphorylation; Polymorphism.
FT DOMAIN 1 39
FT TRANSMEM 40 67
FT DOMAIN 68 77
FT TRANSMEM 78 98
FT DOMAIN 99 111
FT TRANSMEM 112 133
FT DOMAIN 134 150
FT TRANSMEM 151 175
FT DOMAIN 176 206
FT TRANSMEM 207 226
FT DOMAIN 227 242
FT TRANSMEM 243 267
FT DOMAIN 268 284
FT TRANSMEM 285 308
FT DOMAIN 309 360
FT CARBOHYD 183 183
FT CARBOHYD 194 194
FT DISULFID 110 187
FT VARIANT 130 130
FT VARIANT 178 178
FT VARIANT 178 178
SQ SEQUENCE 360 AA; 41402 MW; 51EBE12AD1FAFABF CRC64;
Query Match 17.2%; Score 322.5; DB 1; Length 360;
Best Local Similarity 28.3%; Pred. No. 4.3e-16;
Matches 93; Conservative 58; Mismatches 131; Indels 47; Gaps 9;
QY 29 KIFURSLFYTALLVFSALGNILALCLTCQSKRKINCTGIYVHLAVSDLLFTVALP--GR 86
DB ELFLPPL-YSLVVFVGLLGNVVLVLFYKRLSRMTDVLNLAISDLLFVSLFPWG- 96
QY 87 VVCYVLGSSNPFKGLCLRLTAFLVLYTDTYGVYLMACVSDHYPAVCAHWGRLRTAGR 146
DB 97 ---YAADQWVFGGLCKMISWMLVGFYSGIFVWLMISDRYLAIVHAFSLRARTLY 153
QY 147 ARLVCAVIAITVLVLTQWPLLLMPMTKPLVGLKLAEMEYSSMESVLGLPLMVLVAFVAFGFCG 206
DB 154 GVITSLATWSAVFASLPGLFSTCTVTRNHTYCKTKYSLNSTTWKVLSSLEINILGLVI 213
QY 207 PVGIILSCYMKI--TWKLCSTAREDPVTSRKGRRHRCGLLTLLMLVAVVVCF---SPYH 260
DB 214 PLGMFLCYFSMINTLQHCNKKKNAVK-----MIFAVVVFLGFWTPT- 258
QY 261 LNIKQFVAR--GMLHLPSCAERRAFLLSLOATVALMMNCGITPIIYFFASTHYRKWLL- 317
DB 259 -NIVLLETLELVLDVQDCITFERVLDVYAIQATETLAFVHCLNPLIIVFFLGEKFKYILQ 317
QY 318 -----GILKLGSSSSSSSS 331
DB 318 LFKTCRGLFVLCQYCGLLQIYSADTPSS 346

Search completed: February 19, 2004, 19:33:29
Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 19:29:05 ; Search time 41 Seconds

(without alignments)
2265.828 Million cell updates/sec

Title: US-10-041-615-34

Perfect score: 1876
Sequence: 1 MKLGPYFPTPKTKTKID.....SETPSITQARGSMFLAEHV 360

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum Match 100%

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.23.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_virus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	387.5	20.7	339	4 Q8N5S7	Q8N5S7 homo sapien
2	332.5	17.7	361	11 Q8VHP3	Q8VHP3 cavia porce
3	324	17.3	355	4 Q8YX5	Q8YX5 homo sapien
4	323	17.2	337	4 Q8V19	Q8V19 homo sapien
5	316.5	16.9	372	4 Q9H10	Q9H10 homo sapien
6	314.5	16.8	360	11 Q91ZH4	Q91ZH4 rattus norv
7	313.5	16.7	355	6 Q8HZN4	Q8HZN4 cercopithe
8	310.5	16.6	355	6 Q8HZN5	Q8HZN5 macaca mula
9	309.5	16.5	355	6 Q8HZN6	Q8HZN6 pongo pygma
10	309	16.5	355	6 Q8HZN7	Q8HZN7 callithrix
11	308.5	16.4	360	6 Q8HJW8	Q8HJW8 canis fami
12	306.5	16.3	355	4 Q8I0Z1	Q8I0Z1 homo sapien
13	306.5	16.3	355	6 Q8HZN8	Q8HZN8 pan troglod
14	306.5	16.3	362	4 Q8NE10	Q8NE10 homo sapien
15	305	16.3	352	11 Q8BY68	Q8BY68 mus musculu
16	304.5	16.2	355	6 Q8HZN7	Q8HZN7 gorilla gor

17	304	16.2	342	13 Q93239	O93239 cyprinus ca
18	304	16.2	374	13 Q57466	Q57466 meleagris g
19	303.5	16.2	344	11 Q8BMC0	Q8BMC0 mus musculu
20	300.5	16.0	355	6 Q8HZN3	Q8HZN3 papio hamad
21	300	16.0	367	11 Q8QW6	Q8QW6 mus musculu
22	299.5	16.0	357	13 Q9DE05	Q9DE05 raja erinac
23	296	15.8	328	11 Q8ERK9	Q8ERK9 mus musculu
24	294.5	15.7	331	11 Q8K126	Q8K126 mus musculu
25	294.5	15.7	367	11 Q9J1I9	Q9J1I9 rattus norv
26	293.5	15.6	370	11 Q8BKK1	Q8BKK1 mus musculu
27	290.5	15.5	362	11 Q9JLZ0	Q9JLZ0 rattus norv
28	290.5	15.5	362	11 Q9JW10	Q9JW10 mus musculu
29	289.5	15.4	309	11 Q8S28	Q8S28 mus musculu
30	289.5	15.4	370	11 Q8BLG2	Q8BLG2 mus musculu
31	288.5	15.4	350	4 Q8N6T6	Q8N6T6 homo sapien
32	288.5	15.4	375	11 Q8BY11	Q8BY11 mus musculu
33	288	15.4	365	11 Q8BUD0	Q8BUD0 mus musculu
34	288	15.4	372	13 Q9J237	Q9J237 cyprinus ca
35	287	15.3	346	4 Q8XCO	Q8XCO homo sapien
36	287	15.3	360	13 Q8QFR6	Q8QFR6 xenopus lae
37	286.5	15.3	353	13 Q9PTF7	Q9PTF7 brachydanio
38	286	15.2	377	13 Q98U14	Q98U14 brachydanio
39	284.5	15.2	359	11 Q8K3M7	Q8K3M7 mus musculu
40	283	15.1	343	11 Q8C131	Q8C131 mus musculu
41	283	15.1	361	13 Q90X57	Q90X57 xenopus lae
42	283	15.1	377	11 Q8BVF1	Q8BVF1 mus musculu
43	282.5	15.1	359	11 Q8BHB8	Q8BHB8 mus musculu
44	281	15.0	354	11 Q8CBJ0	Q8CBJ0 mus musculu
45	281	15.0	359	13 Q9PVY7	Q9PVY7 anguilla an

ALIGNMENTS

RESULT 1

Q8N5S7 PRELIMINARY; PRT; 339 AA.
AC Q8N5S7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE G protein-coupled receptor 17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and lung;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC031653; AH31653.1; -
DR InterPro; IPR000276; GPCR_Rhodspan.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCR_Rhodopsin.
DR PROSITE; PS00237; G-PROTEIN RECEPT_F1_1; 1.
DR PROSITE; PS00262; G-PROTEIN RECEPT_F1_2; 1.
KW Receptor.
SQ SEQUENCE 339 AA; 37860 MW; BB0CAFD0FC371D63 CRC64;

Query Match 20.7%; Score 387.5; DB 4; Length 339;
Best Local Similarity 30.2%; Pred. No. 1.3e-27;
Matches 94; Conservative 57; Mismatches 143; Indels 17; Gaps 5;
Qy 36 FYTALVFSALGNIALCLCTQKSRKINCTGTYLVHLAVSDLLFTVALPQVYVYLGSS 95
Db 37 FYLLDFILAVGNTLWFFPIRDHSGTPANVFLHVLAVDSLVLPLRLVYHFSGNH 96
Qy 96 WPGKGLCRITAFVLYTDTYGGVYLMACVSDHYPAVCAHWGPRLRRTAGRLVCVAIW 155
Db 97 WPGFETACRLTGFLFVLYNMVASYFYFTCSADRFALVHVPVKSRLKRRPDIYAHLCALFW 156
Qy 156 TLVLLQTMPLLLMPTKPLVGLKACME-YSSMBSVLGLPLMLVLVAFVAFGFCGPGVILSC 214

```
Db 157 VVAVAMAPLLVSPQVQTNHTVWCLQYREKAS-----HVALVSLAVAFPTTITVTC 211
Qy 215 YMKITWKGSTAREDPVTSRKGHRWGCLLTLLMLVAVVVCFSVPHLNKQFMARGMLHL 274
Db 212 YLLIIRSLRQGRVE-----KELTKAVAMIAVLAIPLVCFVPHVNRVSVVLHYRSHG 266
Qy 275 PSCAERRAFLLSLQATVALMNNCGITPIIYFFASTHYRKMGLGL---KLKG---SSSS 328
Db 267 ASCATQRIILANRITSCUTSUNGALDPIYFFVAEKFHALCNLCOKRLKGGPPSPFEG 326
Qy 329 SSSSSSTPGKA 339
Db 327 KTNESLSAKS 337
```

RESULT 2

```
Q8VHP3 PRELIMINARY; PRT; 361 AA.
AC Q8VHP3;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE CC-chemokine receptor 4.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RA Jopling L.A., Sabroe I., Andrew D.P., Mitchell T.J., Li Y.,
RA Hodge M.R., Williams T.J., Pease J.E.;
RT "The identification, characterization and distribution of guinea pig
RT CCR4 and epitope mapping of a blocking antibody."
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF431971; AAL57488.1;
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS; PRO0237; GPCRHOPOPSN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.
KW Receptor.
SQ SEQUENCE 361 AA; 41064 MW; 9304E897D4FD6839 CRC64;
```

Query Match 17.7%; Score 332.5; DB 11; Length 361;
Best Local Similarity 27.1%; Pred. No. 1.7e-22;
Matches 93; Conservative 67; Mismatches 136; Indels 47; Gaps 8;

```
Qy 15 KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQSKINCTGTYLVHLAV 74
Db 32 KTVKAFG-----ELFLPL-YSLVFLGLGNSVVLVFLPKYKLRSMVDVLLNLAI 84
Qy 75 SLLFTVALP--GRVVCYVLGSSWPFKGKLCRLTAFLVLYDTYGGVLMACVSDHYPAV 132
Db 85 SLLFTVLSLPFWG-----YYAADQWVFGVLCYKMSIWIYGVFGSIFFLVMSIDRYLAI 140
Qy 133 VCAHWGPRLATAGRARLVCAIWTLLVLTQMTPLLMPTKPLVGLKACMEYSMESVGLG 192
Db 141 VHGVFSMRVTFYGVITSLATWAVAFSLPGLLFTSCYTERNHTSCYRYSANSTTWK 200
Qy 193 PLMYLVAFAGFCGPGVGIILSCYKNI--TWKLCSTAREDPVTSRKGHRWGCLLTLLMLV 250
Db 201 VLSSLEINILGLVPLGIMLFCYSMIITRLQHCCKKKKNAVK-----MIF 246
Qy 251 AVVVCF---SPVHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNNCGITPIIYF 306
Db 247 AVVVLFGFTFPYINVLFTLVELEVLCQCSLEKLYDFAQATEFLATHCCLNFIYF 306
Qy 307 FASTHYRKMGLGLK-----LKGSSSSSSSSST 335
Db 307 FLGKFKFYVQLFKTCRGPFPAPQCYCALLRIYSNDTSSSYT 349
```

RESULT 3

```
Q8BYX5 PRELIMINARY; PRT; 355 AA.
AC Q8BYX5;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE CC chemokine receptor 8.
GN CCR8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA DeVries M.E., Xu L., Kelvin A.A., Kelvin D.J.;
RT "Identification and characterization of the CX3CR1 and CCR8
RT promoters."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL: AY016370; AAK08628.1;
DR InterPro: IPR004068; CC_chemkin8.
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS; PRO0530; CHEMOKINER8.
DR PRINTS; PRO0237; GPCRHOPOPSN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
SQ SEQUENCE 355 AA; 40830 MW; D4167F3433B75042 CRC64;
```

Query Match 17.3%; Score 324; DB 4; Length 355;
Best Local Similarity 26.4%; Pred. No. 1e-21;
Matches 89; Conservative 65; Mismatches 145; Indels 38; Gaps 8;

```
Qy 26 QTVKIFLRSLFYTALLVFSALGN--IALCLTCQSKINCTGTYLVHLAVSLLFTVAL 83
Db 31 QTVKGLLAVFYCLLFVLSLGNLSVLVLCVKLRSL--TDVLLNLALSLLFVFSF 88
Qy 84 PGRVVCYVLGSSWPFKGKLCRLTAFLVLYDTYGGVLMACVSDHYPAVCAHWGPRLT 143
Db 89 PFQT--YLLDQWVFGVTCVCKVSGFYIYGFYSMFFITLMSVDRLVAVHVALKVRT 146
Qy 144 AGRARLVCAIWTLLVLTQMTPLLMPTKPLVGLKACMEYSMESVGLGLPLMLVAFAG 203
Db 147 IRMTTICLAVLTAIATIPLLVYQVASEDGLVQCYSFVNOQT---LWKIFTNFKMN 203
Qy 204 FCG---PVGILSCYKMTWKL--CSTAREDPVTSRKGHRWGCLLTLLMLVAVVCFSP 258
Db 204 ILGLLIPTTFMFYCIKILHQLKRCQNHKT-----KAIRLVLIIVIASLLFWVP 253
Qy 259 YHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNNCGITPIIYFFASTHYRKL-- 316
Db 254 FNVVFLTSLHSMILDGCISQQLTYATHVTETISFTHCCVNPVIYAFVGEKFKHLE 313
Qy 317 -----LGILKKGSSSSSSSSSTPGKASS 341
Db 314 IFQKSCQSFNYLGROMPRESCSEKSSSCQCHSRSSS 350
```

RESULT 4

```
Q8IV19 PRELIMINARY; PRT; 337 AA.
AC Q8IV19;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```



```
Query Match 16.8%; Score 314.5; DB 11; Length 360;
Best Local Similarity 28.1%; Pred. No. 7.7e-21;
Matches 93; Conservative 60; Mismatches 133; Indels 45; Gaps 10;

QY 29 KIFLSFYTALVFSALGNILALCTCKSRKINCTGIYVHLAVSDLLFTVALP--GR 86
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db EVLPPL--YSLVFLGLFGNSVVVLFVKYKRLKSNMTDVYLLNLAISDLFLVLSLFFWG- 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 87 VVCYVLGSSWPFQKGLCRITAFVLYTDTYGGVYLMACVSDHYPAVCAHWGPRLR 146
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db ---YAADQWVFGGLGCKIISNMVYGVYSGIFIMLSIDRYLAIVHAFVLSRLRTTY 153
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 147 ARLVCVAITVLVLTQTPMLLMPMTKPLVGLKACMEYSSNESVLGLPLMLVAFAGFCG 206
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db GVTSLTWTSAVFASLDFGLFTCTDENNHCTCKYQVSNSTTKVLSLSLEINVLGLVI 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 207 PVGIILSCYMKI--TWKLCSTAREDPVTSRKGHRWGCLLTLLMLVAVVVCVF-----SPH 260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 PLGIMLFCYSMIITRLRHCK-----NEKNRAVR-----MIFAVVVLFLGFWTPY- 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 261 LNIQOFMAR--GMHLPSCAERRAFLLSLQATVALMMNMGITPIIYFFASTHYRKWL 318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 -NVVLFLETLVEVLQDCTLELYLDVAIOATETLAFIHCLNPNVYFFLGEKFRYIAQ 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 319 ILK-----LKGSSSSSSSSSSST 335
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 318 LFRICRGPVLCQHQDFLQISSADTSSSVT 348
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
Q8HZN4
ID Q8HZN4 PRELIMINARY; PRT; 355 AA.
AC Q8HZN4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Interleukin 8 receptor B CXCR2.
OS Cercopithecus pygerythrus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Cercopithecus.
OX NCBI_TaxID=60710;
RN [1]
RP SEQUENCE FROM N.A.
RA Horlick R.A., Zhao J., Swanson R.N., Webb M.L., Strohl B.,
RA Baldwin J.J., Auld D.S.;
RT "Orthologs of human receptors and methods of use.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF540793; AAN17319.1;
KW Receptor.
SQ SEQUENCE 355 AA; 40381 MW; D464957687BBFBF CRC64;

Query Match 16.7%; Score 313.5; DB 6; Length 355;
Best Local Similarity 28.5%; Pred. No. 9.4e-21;
Matches 95; Conservative 58; Mismatches 153; Indels 27; Gaps 8;

QY 24 ESQTVKIFLSFYTALLVFSALGNILALCTCKSRKINCTGIYVHLAVSDLLFTVAL 83
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 ESLEINKVFVVIYALVFLSLGNSLVMLVILHSRVSITDVYLLNLAIDLFLALT 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 84 PGRVVCYVLGSSWPFQKGLCRITAFVLYTDTYGGVYLMACVSDHYPAVCAHWGPRLR 143
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 97 P--IWAARKNGWIFGTFLCKVSVLLKEVNFYSGILLACISVDYLAIVHA---TRLT 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 144 AGR--ARLVCAITVLVLTQTPMLLMPMTKPLVGLKACMEYSSNESVLGLPLMLVAF 200
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 152 QKRYLVKFCVLSIWSLSLLALPVLPRRTVTVTISPVCYEDMGNTAKRMVLRILPQ 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 201 AIGFCGPVGIILSCY---MKITWKLCSAREDPVTSRKGHRWGCLLTLLMLVAVVVCFS 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 212 TFGFILPLIMLFCYGLTIRLTFK-----AHMGQKRAMRVIFAVVLIIFLLCWL 260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 258 PYHL-NIKQFMARGMHLPSCAERRAFLLSLQATVALMMNMGITPIIYFFASTHYRKWL 316
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 PYHLVLTADTLMTRELNETCQRRNNDQALDATEILGILHSLNPLIYAFIGQKFRHGL 320
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 317 LGILKLGSSSSSS-----SSSTPGKASSET 343
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 LKILATHGLISKDSLPKDSRSPSVGSSSGHT 351
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
Q8HZN6
ID Q8HZN6 PRELIMINARY; PRT; 355 AA.
AC Q8HZN6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Interleukin 8 receptor B CXCR2.
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pongo.
OX NCBI_TaxID=9600;
RN [1]
```

```
Db 261 PYHLVLTADTLMTRELNETCQRRNNDQALDATEILGILHSLNPLIYAFIGQKFRHGL 320
QY 317 LGILKLGSSSSSS-----SSSTPGKASSETPS 345
Db 321 LKILATHGLISKDSLPKDSRSPSVGSSSGHTSS 353

RESULT 8
Q8HZN5
ID Q8HZN5 PRELIMINARY; PRT; 355 AA.
AC Q8HZN5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Interleukin 8 receptor B CXCR2.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RA Horlick R.A., Zhao J., Swanson R.N., Webb M.L., Strohl B.,
RA Baldwin J.J., Auld D.S.;
RT "Orthologs of human receptors and methods of use.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF540792; AAN17318.1;
KW Receptor.
SQ SEQUENCE 355 AA; 40293 MW; 6A9BF9A921F470C5 CRC64;

Query Match 16.6%; Score 310.5; DB 6; Length 355;
Best Local Similarity 26.7%; Pred. No. 1.8e-20;
Matches 95; Conservative 58; Mismatches 151; Indels 27; Gaps 8;

QY 24 ESQTVKIFLSFYTALLVFSALGNILALCTCKSRKINCTGIYVHLAVSDLLFTVAL 83
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 ESLEINKVFVVIYALVFLSLGNSLVMLVILHSRVSITDVYLLNLAIDLFLALT 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 84 PGRVVCYVLGSSWPFQKGLCRITAFVLYTDTYGGVYLMACVSDHYPAVCAHWGPRLR 143
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 97 PIWAAKNGV--WIFGTFLCKVSVLLKEVNFYSGILLACISVDYLAIVHA---TRLT 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 144 AGR--ARLVCAITVLVLTQTPMLLMPMTKPLVGLKACMEYSSNESVLGLPLMLVAF 200
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 152 QKRYLVKFCVLSIWSLSLLALPVLPRRTVTVTISPVCYEDMGNTAKRMVLRILPQ 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 201 AIGFCGPVGIILSCY---MKITWKLCSAREDPVTSRKGHRWGCLLTLLMLVAVVVCFS 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 212 TFGFILPLIMLFCYGLTIRLTFK-----AHMGQKRAMRVIFAVVLIIFLLCWL 260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 258 PYHL-NIKQFMARGMHLPSCAERRAFLLSLQATVALMMNMGITPIIYFFASTHYRKWL 316
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 PYHLVLTADTLMTRELNETCQRRNNDQALDATEILGILHSLNPLIYAFIGQKFRHGL 320
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 317 LGILKLGSSSSSS-----SSSTPGKASSET 343
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 LKILATHGLISKDSLPKDSRSPSVGSSSGHT 351
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
Q8HZN6
ID Q8HZN6 PRELIMINARY; PRT; 355 AA.
AC Q8HZN6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Interleukin 8 receptor B CXCR2.
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pongo.
OX NCBI_TaxID=9600;
RN [1]
```



```
QY 261 LNIKOFMAR--GMLHLPSCARRAFLLSLOATVALMNMNCGITPIIYFFASTHYRKMVLG 318
D 259 -NWVLFLETLVEVLQDCTFRHLDYATQATETLAFVHCCLNPVIYFFLGEKPKKILVQ 317
QY 319 ILK-----LKGSSSSSSSSSSST 335
D 318 LFKTCRGPHWLCQYCHLLQWSPDPSSSYT 348

RESULT 12
Q8IUZ1 PRELIMINARY; PRT; 355 AA.
AC Q8IUZ1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Interleukin 8 receptor, beta.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Brain;
RA Strausberg R.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC037961; AAH37961.1; -.
KW Receptor.
SQ SEQUENCE 355 AA; 40122 MW; 809AB63ADAGA42C6 CRC64;

Query Match 16.3%; Score 306.5; DB 4; Length 355;
Best Local Similarity 28.0%; Pred. No. 4.2e-20;
Matches 94; Conservative 61; Mismatches 144; Indels 37; Gaps 9;

QY 24 ESQTVKIFLSRYTALLVPSALGNILALCLTCQSKRKNCTGIYLVHLAVSDLLFTVAL 83
D 37 ESLEINKYFVVIYALVFLSLGNSLVMLVILYSRGSRVTDVYLLNLALADLLFALT 96
QY 84 PGRVVCYVLGSSWPFPGKGLCLRTAFVLYDTYGGVYLMACVSDHYPAVCAHWGPRLT 143
D 97 PIWAASKVNG--WIFGTFCKVSVLLKEVNFYSGILLACISVDYRLAIVHA---TRT 151
QY 144 AGR--ARLVCAVAIWTLVLTQMPILLMPMT-----KPLVGLKACMEYSMESVGLPLM 195
D 152 QKRYLVKFCISIWGLSLLLALPVLFRRTVYGSNVSP-----ACYEDMGNTANWRMLL 206
QY 196 VLVAFAIGFCGPGVGIILSCY---MKITWKLCTAREDPVTSRGRHWRCGLTLLMLVAV 252
D 207 RILPQSGFIVPLLMILFCYGTILRTLFK-----AHMGQKRAMRVIFAVVLIF 255
QY 253 VVCFSPYHL-NIKOFMARGMLHLPSCARRAFLLSLOATVALMNMNCGITPIIYFFASTH 311
D 256 LLCWLPYNVLADLTMRTOVQTCERRNHIDRALDATEILGILHSLNPLIYAFIGQK 315

RESULT 13
Q8HZN8 PRELIMINARY; PRT; 355 AA.
AC Q8HZN8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Interleukin 8 receptor B CXCR2.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RA Horlick R.A., Zhao J., Swanson R.N., Webb M.L., Strohl B.,
RA Baldwin J.J., Auld D.S.;
RT "Orthologs of human receptors and methods of use.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF540789; AAI17315.1; -.
KW Receptor.
SQ SEQUENCE 355 AA; 40122 MW; 809AB63ADAGA42C6 CRC64;

Query Match 16.3%; Score 306.5; DB 6; Length 355;
Best Local Similarity 28.0%; Pred. No. 4.2e-20;
Matches 94; Conservative 61; Mismatches 144; Indels 37; Gaps 9;

QY 24 ESQTVKIFLSRYTALLVPSALGNILALCLTCQSKRKNCTGIYLVHLAVSDLLFTVAL 83
D 37 ESLEINKYFVVIYALVFLSLGNSLVMLVILYSRGSRVTDVYLLNLALADLLFALT 96
QY 84 PGRVVCYVLGSSWPFPGKGLCLRTAFVLYDTYGGVYLMACVSDHYPAVCAHWGPRLT 143
D 97 PIWAASKVNG--WIFGTFCKVSVLLKEVNFYSGILLACISVDYRLAIVHA---TRT 151
QY 144 AGR--ARLVCAVAIWTLVLTQMPILLMPMT-----KPLVGLKACMEYSMESVGLPLM 195
D 152 QKRYLVKFCISIWGLSLLLALPVLFRRTVYGSNVSP-----ACYEDMGNTANWRMLL 206
QY 196 VLVAFAIGFCGPGVGIILSCY---MKITWKLCTAREDPVTSRGRHWRCGLTLLMLVAV 252
D 207 RILPQSGFIVPLLMILFCYGTILRTLFK-----AHMGQKRAMRVIFAVVLIF 255
QY 253 VVCFSPYHL-NIKOFMARGMLHLPSCARRAFLLSLOATVALMNMNCGITPIIYFFASTH 311
D 256 LLCWLPYNVLADLTMRTOVQTCERRNHIDRALDATEILGILHSLNPLIYAFIGQK 315

RESULT 14
Q8NE10 PRELIMINARY; PRT; 362 AA.
AC Q8NE10;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Testis;
RA Strausberg R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC036661; AAH36661.1; -.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHOODPSN.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR F1_1; 1.
DR PROSITE; PS00262; G-PROTEIN RECEPTOR F1_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 362 AA; 41478 MW; BOC0EC1AE3F81FB8 CRC64;

Query Match 16.3%; Score 306.5; DB 4; Length 362;
Best Local Similarity 23.9%; Pred. No. 4.3e-20;
Matches 79; Conservative 75; Mismatches 148; Indels 29; Gaps 6;

QY 22 NVESQTVKIFLSRYTALLVPSALGNILALCLTCQSKRKNCTGIYLVHLAVSDLLFTV 81
D 36 NNPENKSVLYLSFYIFIFVIGMIANSVWVWNIQAKTGYDTHCYILNLAIDLWVVL 95
QY 82 ALPGRVVCYVYLGSSWPFPGKGLCLRTAFVLYDTYGGVYLMACVSDHYPAVCAHWGPR 141
D [1]
```

Job time : 43 secs

Db 96 TIPVWVSVLQNHQNMGBELTCKVTHLIFSINLFGSIFFLTCMSVDRVLSITYTNTFSS 155
QY 142 RTAGRARLVCAVATWLVLLQTMF--LLMPMTKPLVGLKACM-----EYSSMESVGLGPLM 195
Db 156 RKKWRRVVCILVWLLAFVCSLPDVTYLLKTVTSASNNETYCRSFYPSHSIKEWLIG--M 212
QY 196 VLVAFAGFCGPGVGIILSCYMKITWKLCSAREDPVTSRKGHWEGCLLTLMLVAVVC 255
Db 213 ELVSVVVGFAVFPFSIVAVFYLLARASISSQDQKHSRK-----IIFSYYVVFLVC 264
QY 256 FSPYHLNIRKQFVARGMLHLP-SCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRK 314
Db 265 WLPYHVAVLLDIFSILHYIPFCRLEHALFTALHVTQCLSLVHCCVNPVLSYFINENRY 324
QY 315 WLL-----GILKLGSSSSSSSSSS 334
Db 325 ELMKAFIKYSAKTGLTKLIDASRVSEYTS 355

RESULT 15

Q8BY68 PRELIMINARY; PRT; 352 AA.
AC Q8BY68;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Similar to putative purinergic receptor P2Y10.
CS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; Pubmed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573 (2002).
DR EMBL; AK041740; BAC31049.1; -;
SQ SEQUENCE 352 AA; 40382 MW; C9977E40202AB4EA CRC64;

Query Match 16.3%; Score 305; DB 11; Length 352;
Best Local Similarity 27.9%; Pred. No. 5.7e-20;
Matches 84; Conservative 59; Mismatches 130; Indels 28; Gaps 9;
QY 27 TVKIFLSLP---YTALLVFSALGNILALCLCTCKSRKINCTGIVLVHLSVLDLFTVAL 83
Db 16 TYLTFQYSLYATTYIFIFPGLLANSVALVLCRFISKKNKALIFMINLSVADLAHLSL 75
QY 84 PGRVVCYVLGSSWPFPGKGLCLRTAFVLYTDTYGGVYLMACVSVVDHYPAVV---CAHWGP 139
Db 76 PLRIY-YIYNRWPFQFALCLCFYLKYNMYASIFFLTCTISLQRCLEFLKPPRARNWKR 134
QY 140 RLRTAGRARLVCAVATWLVLLQTMFLLMPMTKPLVGLKAC-----MEYSSMESVGLPL 194
Db 135 RYDVA-----ISAAVWIIVGTACLPLFLRSAGLANNSESCFADLGLQDITWASSIG--- 186
QY 195 MVLVAFAGFCGPGVGIILSCYMKITWKLCSARE--DPVTSRKGHWEGCLLTLMLVAV 252
Db 187 MVTAEELGGVLPVVIITYC-----TWKRSKLEFQDPQPNQIKERK-KALRWVLMCAVVF 241
QY 253 VCFSPYHLNIRKQFVARGMLHLPSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHY 312
Db 242 IVCFTPYHLNIRKQFVARGMLHLPSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHY 301
QY 313 R 313
Db 302 R 302

Search completed: February 19, 2004, 19:34:23

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 19:33:35 ; Search time 43 Seconds
(without alignments)
1328.874 Million cell updates/sec

Title: US-10-041-615-34

Perfect score: 360

Sequence: 1 MTKLGPYFPPTKTKTKID.....SETPSITQARGSMFLAEHW 360

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 10

Total number of hits satisfying chosen parameters: 692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq 19Jun03:*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	360	100.0	360	ABG32286	Human GPCR16 prote
2	86	23.9	110	ABG66698	Human novel polype
3	86	23.9	377	ABG14682	Novel human diagno
4	70	19.4	177	AAU19222	Human G protein-co
5	51	14.2	177	AAU19222	Human G protein-co
6	13	3.6	311	ABG67853	Drosophila melanog
7	12	3.3	106	AAU32649	Novel human secret
8	12	3.3	1262	ABB70839	Drosophila melanog
9	11	3.1	85	AAU32705	Novel human secret

10	11	3.1	86	23	ABP10952	Human ORFX protein
11	11	3.1	94	22	AAU16320	Human novel secret
12	11	3.1	94	24	ABU55389	Human novel polype
13	11	3.1	104	22	ABG58455	Human liver peptid
14	11	3.1	104	22	ABG43053	Peptide #10559 enc
15	11	3.1	104	22	ABG26224	Protein #8223 enco
16	11	3.1	104	22	AAU63955	Human brain expres
17	11	3.1	104	22	AAU76775	Human bone marrow
18	11	3.1	104	22	AAU20996	Peptide #7430 enco
19	11	3.1	104	22	AAU36882	Peptide #10919 enc
20	11	3.1	104	22	ABG45940	Human novel secret
21	11	3.1	112	22	AAU15859	Human novel polype
22	11	3.1	112	24	ABU54928	Human novel polype
23	11	3.1	115	22	AAU32633	Novel human secret
24	11	3.1	115	22	AAU32633	Human polypeptide
25	11	3.1	115	22	AAU00547	Human polypeptide
26	11	3.1	118	22	AAU00737	Novel human secret
27	11	3.1	118	22	AAU32634	C glutamicum prote
28	11	3.1	134	22	AAU91288	Drosophila melanog
29	11	3.1	141	22	ABG69191	Human diagnostic a
30	11	3.1	142	22	AAU19607	Human DITHP polype
31	11	3.1	142	23	ABG60206	Novel human diagno
32	11	3.1	145	22	ABG18589	Novel human diagno
33	11	3.1	145	22	ABG18766	Novel human diagno
34	11	3.1	145	22	ABG19509	Novel human diagno
35	11	3.1	147	21	AAU08159	Arabidopsis thalia
36	11	3.1	148	21	AAU3674	Arabidopsis thalia
37	11	3.1	178	21	AAU08048	Arabidopsis thalia
38	11	3.1	178	21	AAU51989	Arabidopsis thalia
39	11	3.1	186	21	AAU08168	Arabidopsis thalia
40	11	3.1	187	21	AAU43673	Arabidopsis thalia
41	11	3.1	187	22	AAU05009	Cryptosporidium pa
42	11	3.1	188	22	AAU02457	Human polypeptide
43	11	3.1	199	22	AAU16338	Human novel secret
44	11	3.1	199	24	ABU55407	Human novel polype
45	11	3.1	201	22	AAU15883	Human novel secret
			201	24	ABU54952	Human novel polype

ALIGNMENTS

RESULT 1

ABG32286
ID ABG32286 standard; Protein; 360 AA.

AC ABG32286;

DT 15-NOV-2002 (first entry)

DE Human GPCR16 protein.

KW G protein coupled receptor related protein; human; GPCR;
KW cardiomyopathy; atherosclerosis; diabetes; cancer; stroke;
KW Von Hippel-Lindau syndrome; Alzheimer's disease; tuberosus sclerosis;
KW hypercalcaemia; Parkinson's disease; Huntington's disease;
KW cerebral palsy; epilepsy; Lesch-Nyhan syndrome; multiple sclerosis;
KW ataxia-telangiectasia; leukodystrophy; addiction; anxiety; depression;
KW pain; obesity; Crohn's disease; osteoporosis; haemophilia; asthma;
KW inflammatory bowel disease; infertility; hypertension; scleroderma;
KW arthritis; human immunodeficiency virus; autoimmune disease; HIV;
KW infection; graft-versus-host disease.

OS Homo sapiens.

PN WO200264793-A2.

PD 22-AUG-2002.

PF 03-JAN-2002; 2002WO-US000056.

PR 03-JAN-2001; 2001US-259552P.

PR 09-JAN-2001; 2001US-260544P.

PR 20-MAR-2001; 2001US-277405P.


```
SQ Sequence 110 AA;
Query Match 23.9%; Score 86; DB 23; Length 110;
Best Local Similarity 100.0%; Pred. No. 2.8e-75;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 LRTAGARLVCAIWTLLVLTQMPILLMPMTKPLVGLKACMEYSSMESVGLGLPLMLVAF 200
DB 21 LRTAGARLVCAIWTLLVLTQMPILLMPMTKPLVGLKACMEYSSMESVGLGLPLMLVAF 80

QY 201 AIGFCGPVGIILSCYMKITWKLCSTA 226
DB 81 AIGFCGPVGIILSCYMKITWKLCSTA 106

RESULT 3
ABG14682
ID ABG14682 standard; Protein; 377 AA.
XX
AC ABG14682;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #14673.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSEQ-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR N-PSDB; AAS78869.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 45041; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
```

```
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 377 AA;
Query Match 23.9%; Score 86; DB 22; Length 377;
Best Local Similarity 100.0%; Pred. No. 8.8e-75;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 LRTAGARLVCAIWTLLVLTQMPILLMPMTKPLVGLKACMEYSSMESVGLGLPLMLVAF 200
DB 116 LRTAGARLVCAIWTLLVLTQMPILLMPMTKPLVGLKACMEYSSMESVGLGLPLMLVAF 175

QY 201 AIGFCGPVGIILSCYMKITWKLCSTA 226
DB 176 AIGFCGPVGIILSCYMKITWKLCSTA 201

RESULT 4
AAU19222
ID AAU19222 standard; Protein; 177 AA.
XX
AC AAU19222;
XX
DT 04-DEC-2001 (first entry)
XX
DE Human G protein-coupled receptor nGPCR-2343.
XX
KW Human; G protein-coupled receptor; nGPCR-x; antiviral; analgesic;
KW cytotatic; cardiant; antidiabetic; anorectic; hypotensive; hypertensive;
KW antiparkinsonian; nootropic; neuroprotective; antidepressant;
KW viral infection; HIV-1; human immunodeficiency virus; HIV-2; pain;
KW cancer; metabolic disease; cardiovascular disease; type 2 diabetes;
KW obesity; anorexia; hypotension; hypertension; myocardial infarction;
KW atherosclerosis; Parkinson's disease; psychosis; neurological disorder;
KW schizophrenia; migraine; major depression; anxiety; mental disorder;
KW manic depression; dyskinesia; Huntington's disease; Tourette's Syndrome.
XX
OS Homo sapiens.
XX
PN WO200166750-A2.
XX
PD 13-SEP-2001.
XX
PF 08-MAR-2001; 2001WO-US07322.
XX
PR 08-MAR-2000; 2000US-0187581.
XX
PR 08-MAR-2000; 2000US-0187582.
XX
PR 08-MAR-2000; 2000US-0187714.
XX
PR 08-MAR-2000; 2000US-0187715.
XX
PR 08-MAR-2000; 2000US-0187825.
XX
PR 08-MAR-2000; 2000US-0187828.
XX
PR 08-MAR-2000; 2000US-0187829.
XX
PR 08-MAR-2000; 2000US-0187830.
XX
PR 08-MAR-2000; 2000US-0187833.
XX
PR 08-MAR-2000; 2000US-0187874.
XX
PR 08-MAR-2000; 2000US-0187930.
XX
PR 08-MAR-2000; 2000US-0188049.
XX
PR 08-MAR-2000; 2000US-0189294.
XX
PR 08-MAR-2000; 2000US-0187929.
XX
PR 08-MAR-2000; 2000US-0187928.
XX
PA (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI Vogeli G, Wood LS;
XX
DR WPI; 2001-536778/59.
XX
DR N-PSDB; AAS30791.
XX
PT Isolated nucleic acid molecules encoding G protein-coupled receptors
PT termed nGPCR-x, useful in the treatment and diagnosis of viral
PT infections, cancers and mental disorders (e.g. Parkinson's disease and
PT schizophrenia) -
XX
```

PS Claim 31; Page 281; 336pp; English.

XX The invention relates to novel isolated nucleic acid molecules encoding

CC G protein-coupled receptors termed nGPCR-x. nGPCR-x polynucleotides,

CC polypeptides, and modulators may be used in the treatment of diseases and

CC conditions such as infections, such as viral infections caused by HIV-1

CC (human immunodeficiency virus) or HIV-2, pain, cancers, metabolic and

CC cardiovascular diseases and disorders (e.g., type 2 diabetes, obesity,

CC anorexia, hypotension, hypertension, myocardial infarction,

CC atherosclerosis), Parkinson's disease, and psychotic and

CC neurological disorders, including schizophrenia, migraine, major

CC depression, anxiety, mental disorder, manic depression, and

CC dyskinesias, such as Huntington's disease or Tourette's Syndrome

CC and many other diseases and syndromes listed in the specification.

CC nGPCR-x polynucleotides and polypeptides, as well as nGPCR-x

CC modulators, may also be used in diagnostic assays for such diseases or

CC conditions. The present sequence represents a G protein-coupled

CC receptor of the invention.

XX

SQ Sequence 177 AA;

Query Match 19.4%; Score 70; DB 22; Length 177;

Best Local Similarity 100.0%; Pred. No. 1.5e-59;

Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 SRKGRWRGCLLTLLMLVAVVCFSPYHNLKQFMARGMLHLPSCAERRAFLLSLQATVA 292

DB 1 SRKGRWRGCLLTLLMLVAVVCFSPYHNLKQFMARGMLHLPSCAERRAFLLSLQATVA 60

QY 293 LMMNCGITP 302

DB 61 LMMNCGITP 70

RESULT 5

AAU25588

ID AAU25588 standard; Protein; 177 AA.

XX

AC AAU25588;

XX

DT 19-DEC-2001 (first entry)

XX

DE Human G Protein-Coupled Receptor (GPCR) polypeptide #35.

XX

KW Human; G-protein coupled receptor; GPCR; mental disorder; schizophrenia;

KW attention deficit disorder; anxiety; depression; bipolar disorder;

KW neurological disorder; Huntington's disease; dementia; obesity; anorexia;

KW metabolic disorder; Parkinson's disease; Tourette's syndrome; thrombosis;

KW type 2 diabetes; cardiovascular disorder; myocardial infarction; cancer;

KW cardiomyopathy; atherosclerosis; human immunodeficiency virus; HIV;

KW viral infection; immunostimulant; neuroleptic; nootropic; tranquilizer;

KW antidepressant; anorectic; gene therapy.

OS Homo sapiens.

XX

PN WO200162797-A2.

XX

PD 30-AUG-2001.

XX

PF 23-FEB-2001; 2001WO-US05676.

XX

PR 23-FEB-2000; 2000US-0184247.

PR 23-FEB-2000; 2000US-0184303.

PR 23-FEB-2000; 2000US-0184304.

PR 23-FEB-2000; 2000US-0184305.

PR 23-FEB-2000; 2000US-0184397.

PR 03-MAR-2000; 2000US-0186457.

PR 03-MAR-2000; 2000US-0186810.

PR 09-MAR-2000; 2000US-0188084.

PR 13-MAR-2000; 2000US-0188880.

PR 03-APR-2000; 2000US-0194344.

PR 23-JUN-2000; 2000US-0213861.

PR 11-JUL-2000; 2000US-0217369.

PR 11-JUL-2000; 2000US-0217370.

PR 14-JUL-2000; 2000US-0218337.

PR 20-JUL-2000; 2000US-0218492.

XX (PHAA) PHARMACIA & UPJOHN CO.

XX Vogeli G, Wood LS, Parodi LA, Lind P;

PI WPI; 2001-570628/64.

XX N-PSDB; AAS42840.

DR

XX New isolated nucleic acid encoding a new G-protein coupled receptor

PT polypeptide for detecting receptor modulators that can treat mental

PT disorders, such as schizophrenia, anxiety, depression, or obesity -

XX

PS Claim 35; Page 83; 279pp; English.

XX

CC Sequences AAU25554-AAU25616 represent human G-protein coupled receptor

CC (GPCR) polypeptides of the invention. The proteins and their associated

CC DNA sequences can be used to identify compounds which bind to GPCR

CC polypeptides and in screening for compounds that modulate GPCR activity.

CC By screening a human subject for the presence of mutations in GPCR DNA, a

CC GPCR-related disorder or a genetic predisposition can be diagnosed. The

CC sequences can also be used for treatment and prevention of mental

CC disorders such as schizophrenia, attention deficit disorder, anxiety,

CC depression, dementia and bipolar disorder, neurological disorders such as

CC Huntington's disease, Parkinson's disease and Tourette's syndrome,

CC metabolic disorders such as obesity, anorexia and type 2 diabetes,

CC cardiovascular disorders such as thrombosis, myocardial infarction,

CC cardiomyopathy and atherosclerosis, viral infections caused by HIV and

CC cancers.

XX

SQ Sequence 177 AA;

Query Match 14.2%; Score 51; DB 22; Length 177;

Best Local Similarity 100.0%; Pred. No. 4.4e-41;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 LRTAGRARLYCVVAITWTLVLIQTMTPLLMPTKPLVGLKACMEYSSMESVLG 191

DB 48 LRTAGRARLYCVVAITWTLVLIQTMTPLLMPTKPLVGLKACMEYSSMESVLG 98

RESULT 6

ABB67853

ID ABB67853 standard; Protein; 311 AA.

XX

AC ABB67853;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 30351.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR N-PSDB; ABL11956.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 30351; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 311 AA;
Query Match 3.6%; Score 13; DB 22; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.00062;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 324 GSSSSSSSSSSSTP 336
DB 188 GSSSSSSSSSSSTP 200
RESULT 7
RAU32649
ID AAU32649 standard; Protein; 106 AA.
XX AC AAU32649;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #3140.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
OS Homo sapiens.
XX
PW WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.
PR 26-JAN-2001; 2001US-0770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
PS Claim 20; Page 643; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or

CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 106 AA;
Query Match 3.3%; Score 12; DB 22; Length 106;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 323 KGSSSSSSSSSS 334
DB 65 KGSSSSSSSSSS 76
RESULT 8
ABB70839
ID ABB70839 standard; Protein; 1262 AA.
XX AC ABB70839;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 39309.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PW WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEXE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX N-PSDB; ABL14942.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 39309; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1262 AA;

Query Match 3.3%; Score 12; DB 22; Length 1262;
 Best Local Similarity 100.0%; Pred. No. 0.022;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GSSSSSSSSSSST 335
 |||||
 Db 351 GSSSSSSSSSSST 362

RESULT 9

ID AAU32705 standard; Protein; 85 AA.

XX AC AAU32705;

DT 18-DEC-2001 (first entry)

DE Novel human secreted protein #3196.

XX Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX OS Homo sapiens.

XX PN WO200179449-A2.

XX PD 25-OCT-2001.

XX PF 16-APR-2001; 2001WO-US08656.

XX PR 18-APR-2000; 2000US-0552929.

XX PR 26-JAN-2001; 2001US-0770160.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-611725/70.

XX Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -

XX Claim 20; Page 652; 765pp; English.

XX The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.

XX Sequence 85 AA;

Query Match 3.1%; Score 11; DB 22; Length 85;

Best Local Similarity 100.0%; Pred. No. 0.016;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GSSSSSSSSSSS 334
 |||||
 Db 42 GSSSSSSSSSSS 52

RESULT 10

ID ABP10952

XX ABP10952 standard; Protein; 86 AA.

XX AC ABP10952;

XX DT 24-JUN-2002 (first entry)

XX DE Human ORFX protein sequence SEQ ID NO:21886.

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis.

XX OS Homo sapiens.

XX PN WO200192523-A2.

XX PD 06-DEC-2001.

XX PF 29-MAY-2001; 2001WO-US10836.

XX PR 30-MAY-2000; 2000US-206132P.

XX PR 29-AUG-2000; 2000US-228716P.

XX PA (CURA-) CURAGEN CORP.

XX PI Shimkets RA, Leach MD;

XX WPI; 2002-106308/14.

XX N-PSDB; ABN26704.

XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders -

XX Disclosure; SEQ ID 21886; 1037pp; English.

XX The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.

XX N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 86 AA;

Query Match 3.1%; Score 11; DB 23; Length 86;

Best Local Similarity 100.0%; Pred. No. 0.016; Mismatches 0; Indels 0; Gaps 0;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 324 GSSSSSSSSSS 334
Db 27 GSSSSSSSSSS 37

RESULT 11
AAU16320
ID AAU16320 standard; Protein; 94 AA.
XX AC AAU16320;
XX DT 07-NOV-2001 (first entry)
XX DE Human novel secreted protein, Seq ID 1273.
XX KW Human; immunosuppressive; antiarthritic; antirheumatic;
KW cyrostatic; cardiac; vasotropic; cerebroprotective; neurotropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX OS Homo sapiens.
XX FN WO200155322-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01341.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216847.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225366.
PR 14-AUG-2000; 2000US-0225367.
PR 14-AUG-2000; 2000US-0225368.
PR 14-AUG-2000; 2000US-0225370.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 22-AUG-2000; 2000US-0226379.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251866.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 PI
 XX
 XX WPI; 2001-488783/53.
 DR N-PSDB; AAS26307.
 XX
 PT New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 XX Claim 11; SEQ ID No 1273; 980pp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention.

Query Match 3.1%; Score 11; DB 22; Length 94;
 Best Local Similarity 100.0%; Pred. No. 0.018;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GSSSSSSSSSSSS 334
 |||||
 Db 17 GSSSSSSSSSSSS 27

RESULT 12
 ABUS5389
 ID ABUS5389 standard; Protein; 94 AA.
 XX
 AC ABUS5389;
 XX
 DT 18-MAR-2003 (first entry)
 XX
 DE Human novel polypeptide #476.
 XX
 KW Human; neural disorder; immune system disorder; renal disorder;
 KW muscular disorder; respiratory disease; reproductive disorder;
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
 KW haemostatic; antiarteriosclerotic.
 XX
 OS Homo sapiens.
 XX
 PN US2002132753-A1.
 XX
 PD 19-SEP-2002.
 XX
 PF 17-JAN-2001; 2001US-0764864.
 XX
 PR 31-JAN-2000; 2000US-179065P.
 PR 04-FEB-2000; 2000US-180628P.
 PR 28-JUN-2000; 2000US-214886P.
 PR 07-JUL-2000; 2000US-216647P.
 PR 07-JUL-2000; 2000US-216880P.
 PR 11-JUL-2000; 2000US-217487P.
 PR 11-JUL-2000; 2000US-217496P.
 PR 14-JUL-2000; 2000US-218290P.
 PR 26-JUL-2000; 2000US-220963P.
 PR 26-JUL-2000; 2000US-220964P.
 PR 14-AUG-2000; 2000US-224518P.
 PR 14-AUG-2000; 2000US-224519P.
 PR 14-AUG-2000; 2000US-225267P.
 PR 14-AUG-2000; 2000US-225268P.
 PR 14-AUG-2000; 2000US-225270P.
 PR 14-AUG-2000; 2000US-225447P.
 PR 14-AUG-2000; 2000US-225757P.
 PR 14-AUG-2000; 2000US-225758P.
 PR 22-AUG-2000; 2000US-225758P.
 PR 30-AUG-2000; 2000US-226868P.
 PR 01-SEP-2000; 2000US-228242P.
 PR 01-SEP-2000; 2000US-229287P.
 PR 01-SEP-2000; 2000US-229343P.
 PR 01-SEP-2000; 2000US-229344P.
 PR 01-SEP-2000; 2000US-229345P.
 PR 05-SEP-2000; 2000US-229509P.
 PR 05-SEP-2000; 2000US-229513P.
 PR 08-SEP-2000; 2000US-231413P.
 PR 21-SEP-2000; 2000US-234223P.
 PR 21-SEP-2000; 2000US-234274P.
 PR 25-SEP-2000; 2000US-234997P.
 PR 27-SEP-2000; 2000US-235834P.
 PR 29-SEP-2000; 2000US-236327P.
 PR 29-SEP-2000; 2000US-236367P.
 PR 29-SEP-2000; 2000US-236368P.
 PR 29-SEP-2000; 2000US-236369P.
 PR 29-SEP-2000; 2000US-236370P.
 PR 02-OCT-2000; 2000US-236802P.
 PR 02-OCT-2000; 2000US-237037P.
 PR 02-OCT-2000; 2000US-237038P.
 PR 02-OCT-2000; 2000US-237039P.
 PR 13-OCT-2000; 2000US-237040P.
 PR 13-OCT-2000; 2000US-239355P.
 PR 20-OCT-2000; 2000US-240560P.
 PR 20-OCT-2000; 2000US-241785P.
 PR 20-OCT-2000; 2000US-241809P.
 PR 01-NOV-2000; 2000US-244617P.

```
PR 17-NOV-2000; 2000US-249299P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
XX
XX (ROSE/) ROSEN C A.
XX (RUBE/) RUBEN S M.
XX (BARA/) BARASH S C.
XX
XX Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2003-147444/14.
XX N-PSDB; ABX73648.
XX
XX New polypeptides and nucleic acids, useful in gene therapy for
XX treating, inhibiting or preventing e.g. neural, immune system,
XX muscular, respiratory, reproductive, gastrointestinal, pulmonary,
XX cardiovascular or renal disorders -
XX
XX Claim 11; SEQ ID NO 1273; 402pp; English.
XX
XX The invention relates to human novel polypeptides and their associated
XX polynucleotides. The polypeptides and polynucleotides are useful in gene
XX therapy for treating, inhibiting or preventing neural disorders, immune
XX system disorders (e.g. systemic lupus erythematosus, rheumatoid
XX arthritis and multiple sclerosis), muscular disorders, respiratory
XX diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),
XX reproductive disorders, gastrointestinal disorders, pulmonary disorders,
XX cardiovascular disorders (e.g. congenital heart defects, Ebstein's
XX anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute
XX kidney failure and end-stage renal disease), hyperproliferative disorders
XX (e.g. Hodgkin's disease and leukaemia), inflammatory diseases (e.g.
XX septic shock, buritis and appendicitis), allergic reactions and
XX conditions (e.g. asthma), blood related disorders (e.g. thrombosis,
XX atherosclerosis and myocardial infarction) and cancerous diseases.
XX Sequences ABUS4914-ABUS5699 and ABUS5748 represent human novel
XX polypeptides of the invention.
XX
XX SQ Sequence 94 AA;
XX
XX Query Match 3.1%; Score 11; DB 24; Length 94;
XX Best Local Similarity 100.0%; Pred. No. 0.018;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 324 GSSSSSSSSSS 334
XX | | | | | | | |
XX 17 GSSSSSSSSSS 27
XX
XX RESULT 13
XX ABG58455
XX ID ABG58455 standard; Peptide; 104 AA.
XX
XX AC ABG58455;
XX
XX DT 25-FEB-2003 (first entry)
XX
XX DE Human liver peptide, SEQ ID No 37103.
XX
XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX hypercholesterolaemia; coronary heart disease.
XX
XX OS Homo sapiens.
XX
XX PN WO200157273-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00664.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX
```

```
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488898/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analysing gene expression in human adult liver -
XX
XX Claim 27; SEQ ID No 37103; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SEN) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/ fragments). The probe hybridises at high
XX stringency to a nucleic acid molecule expressed in the human adult
XX liver (I) may be used for predicting, measuring and displaying gene
XX expression in samples derived from human adult liver. The genes
XX identified may be involved in genetic liver diseases such as cirrhosis,
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
XX is associated with coronary heart disease. ABG47348-ABG59930 represent
XX human liver single exon encoded peptides of the invention.
XX Note: The sequence information for this patent does not appear in the
XX printed specification but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX
XX SQ Sequence 104 AA;
XX
XX Query Match 3.1%; Score 11; DB 22; Length 104;
XX Best Local Similarity 100.0%; Pred. No. 0.02;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 325 SSSSSSSSSST 335
XX | | | | | | | |
XX 83 SSSSSSSSSST 93
XX
XX RESULT 14
XX ABB43053
XX ID ABB43053 standard; Peptide; 104 AA.
XX
XX AC ABB43053;
XX
XX DT 04-FEB-2002 (first entry)
XX
XX DE Peptide #10559 encoded by human foetal liver single exon probe.
XX
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
XX OS Homo sapiens.
XX
XX PN WO200157277-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00669.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX
XX OS 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
```

XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX Claim 27; SEQ ID NO 35688; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 104 AA;
Query Match 3.1%; Score 11; DB 22; Length 104;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 325 SSSSSSSSSST 335
||| |||||
Db 83 SSSSSSSSSST 93
RESULT 15
ABB26224
ID ABB26224 standard; Protein; 104 AA.
XX
AC ABB26224;
XX
XX 23-JAN-2002 (first entry)
XX
DE Protein #8223 encoded by probe for measuring heart cell gene expression.
XX Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX
OS Homo sapiens.
XX
XX WO200157274-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00666.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
XX Claim 15; SEQ ID No 27994; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such

CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 104 AA;
Query Match 3.1%; Score 11; DB 22; Length 104;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 325 SSSSSSSSSST 335
||| |||||
Db 83 SSSSSSSSSST 93
Search completed: February 19, 2004, 19:37:18
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 19:36:26 ; Search time 21 Seconds
(without alignments)
725.329 Million cell updates/sec

Title: US-10-041-615-34

Perfect score: 360

Sequence: 1 MKLGPYTPPTKIKTKKD.....SETPSITCARSGMFLAEHV 360

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 10

Total number of hits satisfying chosen parameters: 97

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/1/aaa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/aaa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/aaa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/aaa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/aaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/aaa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	3.1	43	4	US-07-757-022B-12
2	11	3.1	192	4	US-07-757-022B-90
3	11	3.1	204	4	US-07-757-022B-32
4	11	3.1	208	4	US-07-757-022B-132
5	11	3.1	209	4	US-07-757-022B-94
6	11	3.1	217	4	US-07-757-022B-76
7	11	3.1	220	4	US-07-757-022B-96
8	11	3.1	231	4	US-07-757-022B-30
9	11	3.1	296	4	US-07-757-022B-70
10	11	3.1	310	3	US-08-651-136C-22
11	11	3.1	310	4	US-09-229-911A-22
12	11	3.1	372	4	US-07-757-022B-64
13	11	3.1	422	4	US-07-757-022B-68
14	11	3.1	423	4	US-07-757-022B-66
15	11	3.1	463	4	US-07-757-022B-54
16	11	3.1	551	4	US-09-252-991A-27787
17	11	3.1	568	1	US-08-320-553-30
18	11	3.1	568	3	US-08-545-860D-30
19	11	3.1	568	5	PCT-US94-04496-30
20	11	3.1	609	4	US-09-198-452A-579
21	11	3.1	693	4	US-09-107-532A-5812
22	11	3.1	703	3	US-08-910-925-4
23	11	3.1	717	3	US-08-910-925-1
24	11	3.1	743	3	US-08-910-925-3
25	11	3.1	1049	4	US-07-757-022B-58
26	11	3.1	1140	4	US-07-757-022B-104
27	11	3.1	1313	4	US-07-757-022B-142

28 11 3.1 1314 4 US-07-757-022B-50 Sequence 50, Appl
29 11 3.1 1354 4 US-07-757-022B-48 Sequence 48, Appl
30 11 3.1 1363 4 US-07-757-022B-52 Sequence 52, Appl
31 11 3.1 1404 4 US-07-757-022B-2 Sequence 2, Appl
32 11 3.1 1404 4 US-07-757-022B-62 Sequence 62, Appl
33 10 2.8 130 2 US-08-630-822A-90 Sequence 90, Appl
34 10 2.8 130 2 US-09-005-069-90 Sequence 90, Appl
35 10 2.8 130 4 US-09-171-156A-39 Sequence 39, Appl
36 10 2.8 130 4 US-09-004-730A-39 Sequence 39, Appl
37 10 2.8 130 4 US-08-981-799A-39 Sequence 39, Appl
38 10 2.8 148 4 US-09-461-325-453 Sequence 453, Appl
39 10 2.8 218 3 US-09-068-655-7 Sequence 7, Appl
40 10 2.8 223 4 US-09-009-816-4 Sequence 4, Appl
41 10 2.8 229 4 US-09-604-978-9 Sequence 9, Appl
42 10 2.8 228 4 US-09-604-728-9 Sequence 9, Appl
43 10 2.8 349 4 US-09-162-524-3 Sequence 3, Appl
44 10 2.8 367 4 US-08-213-419B-15 Sequence 15, Appl
45 10 2.8 367 4 US-09-009-816-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-07-757-022B-12
; Sequence 12, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: AMINO ACID

```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-12

Query Match      3.1%; Score 11; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
Db 12 SSSSSSSSSST 22

RESULT 2
US-07-757-022B-90
; Sequence 90, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 90:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-90

Query Match      3.1%; Score 11; DB 4; Length 192;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
Db 12 SSSSSSSSSST 22

RESULT 3
US-07-757-022B-92
; Sequence 92, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-92

Query Match      3.1%; Score 11; DB 4; Length 204;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
Db 168 SSSSSSSSSST 178

RESULT 4
US-07-757-022B-132
; Sequence 132, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
```

```

; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 132:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 208 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-132

Query Match 3.1%; Score 11; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
Db 168 SSSSSSSSSST 178

RESULT 5
US-07-757-022B-94
; Sequence 94, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.

```

```

; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 209 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-94

Query Match 3.1%; Score 11; DB 4; Length 209;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
Db 168 SSSSSSSSSST 178

RESULT 6
US-07-757-022B-76
; Sequence 76, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530

```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELEPHONE: (617)876-1170
; TELECOMMUNICATION INFORMATION:
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 217 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-76

Query Match          3.1%, Score 11; DB 4; Length 217;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
DB 128 SSSSSSSSSST 138

RESULT 7
US-07-757-022B-96
; Sequence 96, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Bridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-96

Query Match          3.1%, Score 11; DB 4; Length 220;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
DB 158 SSSSSSSSSST 178

RESULT 8
US-07-757-022B-30
; Sequence 30, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Bridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 220 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-96
```

SEQUENCE CHARACTERISTICS:
LENGTH: 231 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-30

Query Match 3.1%; Score 11; DB 4; Length 231;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
Db 168 SSSSSSSSSST 178

RESULT 9

US-07-757-022B-70
Sequence 70, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/07/757,022B
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA: US 07/546,114
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseir, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)976-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 296 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-70

Query Match 3.1%; Score 11; DB 4; Length 296;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
Db 168 SSSSSSSSSST 178

RESULT 10

US-08-651-136C-22
Sequence 22, Application US/08651136C
Patent No. 6001639
GENERAL INFORMATION:
APPLICANT: Schulein, Martin
APPLICANT: Andersen, Lene N.
APPLICANT: Lassen, Soren F.
APPLICANT: Kauppinen, Markus S.
APPLICANT: Lange, Lene
APPLICANT: Nielsen, Ruby I.
APPLICANT: Ihara, Michiko
APPLICANT: Takagi, Shinobu
TITLE OF INVENTION: No. 6001639el Endoglucanases
NUMBER OF SEQUENCES: 109
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6001639o No. 6001639disk of No. 6001639th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/651,136C
APPLICATION NUMBER: US/08/651,136C
FILING DATE: 21-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4366.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-651-136C-22

Query Match 3.1%; Score 11; DB 3; Length 310;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
Db 248 SSSSSSSSSST 258

RESULT 11

US-09-229-911A-22
Sequence 22, Application US/09229911A
Patent No. 6387690
GENERAL INFORMATION:
APPLICANT: Schulein, Martin
APPLICANT: Andersen, Lene N.
APPLICANT: Lassen, Soren F.
APPLICANT: Kauppinen, Markus S.
APPLICANT: Lange, Lene
APPLICANT: Nielsen, Ruby I.
APPLICANT: Ihara, Michiko

;; Takagi, Shinobu
;; TITLE OF INVENTION: No. 6387690el Endoglucanases
;; NUMBER OF SEQUENCES: 109
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: No. 6387690c No. 6387690disk of No. 6387690th America, Inc.
;; STREET: 405 Lexington Avenue, 64th Floor
;; CITY: New York
;; STATE: New York
;; COUNTRY: United States of America
;; ZIP: 10174-6401
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent in Release #1.0, Version #1.30
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/229,911A
;; FILING DATE: 13-Jan-1993
;; CLASSIFICATION: {UNKNOWN}
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/651,136
;; FILING DATE: 21-MAY-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Lambiris, Elias J.
;; REGISTRATION NUMBER: 33,728
;; REFERENCE/DOCKET NUMBER: 4366.200-US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 212-867-0123
;; TELEFAX: 212-878-9655
;;
;; INFORMATION FOR SEQ ID NO: 22:
;; SEQUENCE DESCRIPTION: SEQ ID NO: 22:
US-09-229-911A-22
;;
Query Match 3.1%; Score 11; DB 4; Length 310;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 325 SSSSSSSSSSST 335
DB 248 SSSSSSSSSSST 258
;;
RESULT 12
US-07-757-022B-64
; Sequence 64, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989

;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/643,502
;; FILING DATE: 18-JAN-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/546,114
;; FILING DATE: 29-JUN-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/457,196
;; FILING DATE: 29-DEC-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/390,901
;; FILING DATE: 08-AUG-1989
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Geert, Luann
;; REGISTRATION NUMBER: 31,822
;; REFERENCE/DOCKET NUMBER: GI 5190
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617)876-1170
;; TELEFAX: (617)876-5851
;; INFORMATION FOR SEQ ID NO: 64:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 372 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-07-757-022B-64
;;
Query Match 3.1%; Score 11; DB 4; Length 372;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 325 SSSSSSSSSSST 335
DB 77 SSSSSSSSSSST 87
;;
RESULT 13
US-07-757-022B-68
; Sequence 68, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-68

Query Match 3.1%; Score 11; DB 4; Length 422;
Best Local Similarity 100.0%; Pred. No. 0.031;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
Db 127 SSSSSSSSSST 137

RESULT 14
US-07-757-022B-66
Sequence 66, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-66

Query Match 3.1%; Score 11; DB 4; Length 423;
Best Local Similarity 100.0%; Pred. No. 0.031;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
Db 128 SSSSSSSSSST 138

RESULT 15
US-07-757-022B-54
Sequence 54, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-54

Query Match 3.1%; Score 11; DB 4; Length 463;
Best Local Similarity 100.0%; Pred. No. 0.033;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 325 SSSSSSSSSST 335

Db 168 SSSSSSSSSST 178

Search completed: February 19, 2004, 19:39:48
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 19:37:21 ; Search time 40 Seconds

(without alignments)
1884.441 Million cell updates/sec

Title: US-10-041-615-34

Perfect score: 360

Sequence: 1 MIKLGPFYPTPTKIKTKIKD.....SETPSITQAGSMFLAEHVY 360

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 801455 seqs, 20932283 residues

Word size : 10

Total number of hits satisfying chosen parameters: 763

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	360	100.0	360	12	US-10-041-615-34
2	110	30.6	229	12	Sequence 34, Appl
3	110	30.6	323	12	Sequence 109, App
4	110	30.6	323	12	Sequence 1620, Ap
5	70	19.4	177	12	Sequence 1294, Ap
6	51	14.2	177	12	Sequence 198, App
7	11	3.1	43	14	Sequence 95, Appl
8	11	3.1	94	10	Sequence 12, Appl
9	11	3.1	104	9	Sequence 1273, Ap
10	11	3.1	112	10	Sequence 41522, A
11	11	3.1	134	10	Sequence 812, App
12	11	3.1	187	12	Sequence 5042, Ap
13	11	3.1	192	14	Sequence 4, Appli
14	11	3.1	199	10	Sequence 90, Appl
15	11	3.1	201	10	Sequence 1291, Ap
					Sequence 636, App

16	11	3.1	204	14	US-10-124-557-92	Sequence 92, Appl
17	11	3.1	208	14	US-10-124-557-132	Sequence 132, App
18	11	3.1	209	14	US-10-124-557-94	Sequence 94, Appl
19	11	3.1	217	14	US-10-124-557-76	Sequence 76, Appl
20	11	3.1	220	14	US-10-124-557-96	Sequence 96, Appl
21	11	3.1	231	14	US-10-124-557-30	Sequence 30, Appl
22	11	3.1	243	11	US-09-533-029-34	Sequence 34, Appl
23	11	3.1	243	11	US-09-533-455-416	Sequence 416, App
24	11	3.1	243	11	US-09-934-455-416	Sequence 2074, Ap
25	11	3.1	243	12	US-10-374-780A-2074	Sequence 3, Appli
26	11	3.1	275	12	US-10-148-687-3	Sequence 3, Appli
27	11	3.1	296	14	US-10-124-557-70	Sequence 70, Appl
28	11	3.1	310	15	US-10-007-521-22	Sequence 22, Appl
29	11	3.1	324	9	US-09-734-417-2	Sequence 2, Appli
30	11	3.1	328	12	US-10-148-687-1	Sequence 1, Appli
31	11	3.1	372	14	US-10-124-557-64	Sequence 64, Appl
32	11	3.1	403	12	US-10-369-493-4127	Sequence 4127, Ap
33	11	3.1	422	14	US-10-124-557-68	Sequence 68, Appl
34	11	3.1	423	14	US-10-124-557-66	Sequence 66, Appl
35	11	3.1	447	12	US-09-833-345-1720	Sequence 1720, Ap
36	11	3.1	448	12	US-10-029-386-32421	Sequence 32421, A
37	11	3.1	448	12	US-10-108-260A-2481	Sequence 2481, A8
38	11	3.1	450	15	US-10-153-668-268	Sequence 268, App
39	11	3.1	450	15	US-10-153-668-380	Sequence 380, App
40	11	3.1	463	14	US-10-124-557-54	Sequence 382, App
41	11	3.1	503	10	US-09-801-368-424	Sequence 54, Appl
42	11	3.1	503	12	US-10-369-493-1943	Sequence 1943, Ap
43	11	3.1	506	12	US-10-369-493-4115	Sequence 4115, Ap
44	11	3.1	544	12	US-10-032-585-7924	Sequence 7924, Ap
45	11	3.1	547	12	US-10-264-049-2651	Sequence 2651, Ap

ALIGNMENTS

RESULT 1

US-10-041-615-34
; Sequence 34, Application US/10041615
; Publication No. US20040014038A1
; GENERAL INFORMATION:
; APPLICANT: Casman, Stacie J
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Ellerman, Karen
; APPLICANT: Smithson, Glennda
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: No. US20040014038A1el GPCR-Like Proteins and Nucleic Acids Encodi
; FILE REFERENCE: 21402-233-061
; CURRENT APPLICATION NUMBER: US/10/041,615
; CURRENT FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: 60/259,552
; PRIOR FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: 60/260,544
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/277,405
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 34
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-041-615-34

Query Match 100.0%; Score 360; DB 12; Length 360;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIKLGPFYPTPTKIKTKIKLNVESQTVKIFLRSFYTALLVFSALGNILALCLTCQKSR 60

Db 1 MIKLGPFYPTPTKIKTKIKLNVESQTVKIFLRSFYTALLVFSALGNILALCLTCQKSR 60

QY 61 KINCTGIVLHVLAUSDLLFTVALPGRVVYVGLSSWPFKGLCLRTAFVLVTDYGGVYL 120

Db 61 KINCTGIVLHVAISDLFTVALPGRVVCYVILGSSWPEKGLCLRTAFVLYTDTYGGVYL 120
QY 121 MACVSDVHPVAVVCAHWGPRLTAGRARLVCVAIWTLVLLQTMPELLLMPMTKPLVGLKAC 180
Db 121 MACVSDVHPVAVVCAHWGPRLTAGRARLVCVAIWTLVLLQTMPELLLMPMTKPLVGLKAC 180
QY 181 MEYSSMESVGLGLPLMLVAFAGCGPVGIIILSCYMKITWKLCTAREDPVTSRKGHR 240
Db 181 MEYSSMESVGLGLPLMLVAFAGCGPVGIIILSCYMKITWKLCTAREDPVTSRKGHR 240
QY 241 GCLLTLLMLVAVVVCSPHYHLNKPQFMARGLHLPSCAERRAFLLSLQATVALMNNCGI 300
Db 241 GCLLTLLMLVAVVVCSPHYHLNKPQFMARGLHLPSCAERRAFLLSLQATVALMNNCGI 300
QY 301 TPIIYFPASTHYKWLGLGLKLGSSSSSSSSSSSTPGKASSETPSITQARGSMFLAHHV 360
Db 301 TPIIYFPASTHYKWLGLGLKLGSSSSSSSSSSSTPGKASSETPSITQARGSMFLAHHV 360

RESULT 2

US-10-041-615-109
; Sequence 109, Application US/10041615
; Publication No. US20040014038A1
; GENERAL INFORMATION:
; APPLICANT: Casman, Stacie J
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Ellerman, Karen
; APPLICANT: Smithson, Glenda
; APPLICANT: Kerkuda, Ramesh
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: NO. US20040014038A1el GPCR-Like Proteins and Nucleic Acids Encodi
; FILE REFERENCE: 21402-233-061
; CURRENT APPLICATION NUMBER: US/10/041,615
; CURRENT FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: 60/259,552
; PRIOR FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: 60/260,544
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/277,405
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: Curaseqlist version 0.1
; SEQ ID NO 109
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-041-615-109

Query Match 30.6%; Score 110; DB 12; Length 229;
Best Local Similarity 100.0%; Pred. No. 5.9e-96;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 117 GVIIMACVSDVHPVAVVCAHWGPRLTAGRARLVCVAIWTLVLLQTMPELLLMPMTKPLVG 176
Db 116 GVIIMACVSDVHPVAVVCAHWGPRLTAGRARLVCVAIWTLVLLQTMPELLLMPMTKPLVG 175
QY 177 KLACMEYSSMESVGLGLPLMLVAFAGCGPVGIIILSCYMKITWKLCTA 226
Db 176 KLACMEYSSMESVGLGLPLMLVAFAGCGPVGIIILSCYMKITWKLCTA 225

RESULT 3

US-10-017-161-1620
; Sequence 1620, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152

; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1620
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-1620

Query Match 30.6%; Score 110; DB 12; Length 323;
Best Local Similarity 100.0%; Pred. No. 8e-96;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 117 GVIIMACVSDVHPVAVVCAHWGPRLTAGRARLVCVAIWTLVLLQTMPELLLMPMTKPLVG 176
Db 58 GVIIMACVSDVHPVAVVCAHWGPRLTAGRARLVCVAIWTLVLLQTMPELLLMPMTKPLVG 117
QY 177 KLACMEYSSMESVGLGLPLMLVAFAGCGPVGIIILSCYMKITWKLCTA 226
Db 118 KLACMEYSSMESVGLGLPLMLVAFAGCGPVGIIILSCYMKITWKLCTA 167

RESULT 4

US-10-292-798-1294
; Sequence 1294, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1294
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-292-798-1294

Query Match 30.6%; Score 110; DB 12; Length 323;
Best Local Similarity 100.0%; Pred. No. 8e-96;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 117 GVIIMACVSDVHPVAVVCAHWGPRLTAGRARLVCVAIWTLVLLQTMPELLLMPMTKPLVG 176
Db 58 GVIIMACVSDVHPVAVVCAHWGPRLTAGRARLVCVAIWTLVLLQTMPELLLMPMTKPLVG 117
QY 177 KLACMEYSSMESVGLGLPLMLVAFAGCGPVGIIILSCYMKITWKLCTA 226
Db 118 KLACMEYSSMESVGLGLPLMLVAFAGCGPVGIIILSCYMKITWKLCTA 167

RESULT 5

US-09-801-944B-198
; Sequence 198, Application US/09801944B
; Publication No. US20040014169A1
; GENERAL INFORMATION:
; APPLICANT: Vogel, Gabriel
; APPLICANT: Wood, Linda S
; TITLE OF INVENTION: NO. US20040014169A1el G Protein-Coupled Receptors
; FILE REFERENCE: 001000U1
; CURRENT APPLICATION NUMBER: US/09/801,944B

```
; CURRENT FILING DATE: 2003-08-01
; PRIOR APPLICATION NUMBER: 60/187,828
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,715
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,929
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,930
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,825
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,833
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,830
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,829
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,582
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/187,581
; PRIOR FILING DATE: 2000-03-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 198
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-801-944B-198

Query Match      19.4%; Score 70; DB 12; Length 177;
Best Local Similarity 100.0%; Pred. No. 4.1e-58;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      233 SRKGRHWRCGLTLLMLVAVVCFSPYHLNKKQFMARGMLHLPSCAERRAFLLSLQATVA 292
Db      1 SRKGRHWRCGLTLLMLVAVVCFSPYHLNKKQFMARGMLHLPSCAERRAFLLSLQATVA 60

QY      293 LMNMGITP 302
Db      61 LMNMGITP 70

RESULT 6
US-09-791-932-95
; Sequence 95, Application US/09791932
; Publication No. US20030003451A1
; GENERAL INFORMATION:
; APPLICANT: Vogeli, Gabriel
; APPLICANT: Parodi, Luis A.
; APPLICANT: Hiebsch, Ronald R.
; APPLICANT: Lind, Peter
; APPLICANT: Kaytes, Paul S.
; APPLICANT: Ruff, Valerie
; APPLICANT: Huff, Rita M.
; APPLICANT: Wood, Linda S.
; TITLE OF INVENTION: No. US20030003451A1el G Protein-Coupled Receptors Cross-Referen
; FILE REFERENCE: 00325 US1
; CURRENT APPLICATION NUMBER: US/09/791,932
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/184,305
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,304
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,303
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,397
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,247
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/189,880
; PRIOR FILING DATE: 2000-03-13
; PRIOR APPLICATION NUMBER: 60/217,369
```

```
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/217,370
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/218,492
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 60/186,810
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: 60/188,064
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: 60/186,457
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: 60/213,861
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/194,344
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/218,337
; PRIOR FILING DATE: 2000-07-14
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 95
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-932-95

Query Match      14.2%; Score 51; DB 11; Length 177;
Best Local Similarity 100.0%; Pred. No. 4.3e-40;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      141 LRTAGRAIVCVVAITWTLTLLQTPMLLMPMTKPLVGLKACMEYSSMESVLG 191
Db      48 LRTAGRAIVCVVAITWTLTLLQTPMLLMPMTKPLVGLKACMEYSSMESVLG 98

RESULT 7
US-10-124-557-12
; Sequence 12, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
```

```

; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-124-557-12

Query Match 3.1%; Score 11; DB 14; Length 43;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
Db 12 SSSSSSSSSST 22
|||||

RESULT 8
US-09-764-864-1273
; Sequence 1273, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1273
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-1273

Query Match 3.1%; Score 11; DB 10; Length 94;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GSSSSSSSSSS 334
Db 17 GSSSSSSSSSS 27
|||||

RESULT 9
US-09-864-761-41522
; Sequence 41522, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aetonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
```

```

; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 41522
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL161747.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.5
; OTHER INFORMATION: EST HUMAN HIT: BF340331.1, EVALUATE 2.00e-24
; OTHER INFORMATION: SWISSPROT HIT: P53236, EVALUATE 1.50e+00
US-09-864-761-41522

Query Match 3.1%; Score 11; DB 9; Length 104;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
Db 83 SSSSSSSSSST 93
|||||

RESULT 10
US-09-764-864-812
; Sequence 812, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 812
; LENGTH: 112
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
US-09-764-864-812

Query Match
Best Local Similarity 3.1%; Score 11; DB 10; Length 112;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 324 GSSSSSSSSSS 334
Db 35 GSSSSSSSSSS 45

RESULT 11
US-09-738-626-5042
; Sequence 5042, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENO, AKIHITO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR FILING DATE: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5042
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5042

Query Match
Best Local Similarity 3.1%; Score 11; DB 10; Length 134;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 325 SSSSSSSSSST 335
Db 42 SSSSSSSSSST 52

RESULT 12
US-10-148-687-4
; Sequence 4, Application US/10148687
; Publication No. US20030185836A1
; GENERAL INFORMATION:
; APPLICANT: WINTER, Gerhard
; APPLICANT: SLADE, Martin Basil
; APPLICANT: WILLIAMS, Keith Leslie
; APPLICANT: GOOLEY, Andrew Arthur
; APPLICANT: Macquarie Research Ltd
; TITLE OF INVENTION: Cryptosporidium sporozoite antigens
; FILE REFERENCE: 047763-5019-US
; CURRENT APPLICATION NUMBER: US/10/148,687
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: PCT/AU00/01492
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: AU PQ4400
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 67

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
US-10-148-687-4

Query Match
Best Local Similarity 3.1%; Score 11; DB 12; Length 187;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 324 GSSSSSSSSSS 334
Db 6 GSSSSSSSSSS 16

RESULT 13
US-10-124-557-90
; Sequence 90, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; APPLICANT: Clark, Stephen C.
; APPLICANT: Jacobs, Kenneth
; APPLICANT: Hewick, Rodney M.
; APPLICANT: Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ceerz, Luan
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 90:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 90:
US-10-124-557-90

Query Match
Best Local Similarity 3.1%; Score 11; DB 14; Length 192;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 325 SSSSSSSSSST 335
| | | | | | | | | |
Db 168 SSSSSSSSSST 178

RESULT 14

US-09-764-864-1291
; Sequence 1291, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1291
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (187)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1291

Query Match 3.1%; Score 11; DB 10; Length 199;
Best Local Similarity 100.0%; Pred. No. 0.041;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GSSSSSSSSSS 334
| | | | | | | | | |
Db 31 GSSSSSSSSSS 41

RESULT 15

US-09-764-864-836
; Sequence 836, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 836
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-836

Query Match 3.1%; Score 11; DB 10; Length 201;
Best Local Similarity 100.0%; Pred. No. 0.042;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GSSSSSSSSSS 334
| | | | | | | | | |
Db 33 GSSSSSSSSSS 43

Search completed: February 19, 2004, 19:40:41
Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 19:35:36 ; Search time 20 Seconds
(without alignments)

1731.036 Million cell updates/sec

Title: US-10-041-615-34

Perfect score: 360

Sequence: 1 MVLGPFPPFKIKTKID.....SETPSIQSGMFLAHVV 360

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 10

Total number of hits satisfying chosen parameters: 218

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_76.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	3.1	135	2 A42098	transcription fact
2	11	3.1	162	2 C85356	glycine-rich prote
3	11	3.1	192	2 S59537	heat shock transcr
4	11	3.1	243	2 T51989	ethylene responsiv
5	11	3.1	262	2 T40941	hypothetical serin
6	11	3.1	284	2 A35419	neutrophil protein
7	11	3.1	288	2 B40722	homeotic protein l
8	11	3.1	306	2 T47832	hypothetical prote
9	11	3.1	323	2 G96544	hypothetical prote
10	11	3.1	416	2 G83656	hypothetical prote
11	11	3.1	446	2 T45525	WSC4 homolog [impo
12	11	3.1	447	2 E83465	conserved hypothet
13	11	3.1	467	2 S30839	UTR2 protein - yea
14	11	3.1	503	2 S63257	probable membrane
15	11	3.1	568	2 T39411	AF-9 protein - hum
16	11	3.1	583	1 S22544	transcription fact
17	11	3.1	641	1 A41932	fibrinogen alpha-I
18	11	3.1	710	2 S28014	OUTD protein - Erw
19	11	3.1	734	2 B42680	nucleolus-cytoplas
20	11	3.1	741	2 T48694	probable transcrip
21	11	3.1	742	2 A49672	transcription fact
22	11	3.1	772	2 A55004	transcription fact
23	11	3.1	779	2 T49717	related to BCS1 pr
24	11	3.1	962	2 S03818	carboxymethylcellu
25	11	3.1	1019	2 T30148	hypothetical prote
26	11	3.1	1052	2 C64221	hypothetical 114K
27	11	3.1	1215	2 I52862	autoantigen - huma
28	11	3.1	1237	2 A34598	ecdysone-induced p
29	11	3.1	1341	2 S50366	probable membrane

30 11 3.1 1390 2 T14004 trfA protein - sli
31 11 3.1 1394 2 B34598 ecdysone-induced p
32 11 3.1 1443 2 S05979 steroid hormone re
33 11 3.1 1723 2 H86557 polymorphic membra
34 11 3.1 1723 2 E72067 polymorphic membra
35 11 3.1 1732 2 C81601 polymorphic membra
36 11 3.1 1823 2 S28974 vitellogenin precu
37 11 3.1 1912 2 T29088 vitellogenin I pre
38 10 2.8 98 2 S13623 ht-en protein - le
39 10 2.8 105 2 C86242 protein T1685.1 [i
40 10 2.8 107 2 B85356 glycine-rich prote
41 10 2.8 109 2 T12208 probable cytochrom
42 10 2.8 125 2 S69871 hypothetical prote
43 10 2.8 135 2 T09876 denydrin - upland
44 10 2.8 145 2 S04042 embryonic abundant
45 10 2.8 149 2 T23179 hypothetical prote

ALIGNMENTS

RESULT 1

A42098

transcription factor Oct-2B, octamer-binding - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 15-Oct-1999

C:Accession: A42098

Rifanaka, M.; Lai, J.S.; Herr, W.

Cell 68, 755-767, 1992

A:Title: Promoter-selective activation domains in Oct-1 and Oct-2 direct differential ac

A:Reference number: A42098; MUID:92154688; PMID:1739980

A>Note: nucleotide sequence is not given

A:Accession: A42098

A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra

A:Molecule type: mRNA

A:Residues: 1-135 <TAN>

C:Superfamily: unassigned homeobox proteins; homeobox homology

C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match

Best Local Similarity 3.1%; Score 11; DB 2; Length 135;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335

DB 103 SSSSSSSSSST 113

RESULT 2

C85356

glycine-rich protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Mar-2001

C:Accession: C85356

R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488; PMID:10617198

A:Accession: C85356

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-162 <STO>

A:Cross-references: GB:NC_001268; NID:G7269947; PIDN:CAB79764.1; GSPDB:GN00140

C:Genetics:

A:Gene: At4g30460

A:Map position: 4

C:Superfamily: glycine-rich cell wall structural protein 1

Query Match

Best Local Similarity 3.1%; Score 11; DB 2; Length 162;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 SSSSSSSSSSS 334

```

Db          69 GSSSSSSSSSS 79
|||||
RESULT 3
S59537
heat shock transcription factor HSP21 - soybean (fragment)
C:Species: Glycine max (soybean)
C>Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 02-Sep-2000
C:Accession: S59537; S52303
R:Carnecka-Verner, E.; Yuan, C.X.; Fox, P.C.; Gurley, W.B.
Plant Mol. Biol. 29, 37-51, 1995
A:Title: Isolation and characterization of six heat shock transcription factor cDNA clones
A:Reference number: S59537; MUID:96017612; PMID:7579166
A:Accession: S59537
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-192 <CZA>
A:Cross-references: EMBL:Z46952
C:Genetics:
A:Gene: HSP21
C:Superfamily: tomato heat shock transcription factor HSP24; HSF DNA-binding domain homeodomain
C:Keywords: DNA binding; heat shock; leucine zipper; nucleus; stress-induced protein; transcription factor
P:46-141/Domain: HSF DNA-binding domain homology <HSF>
Query Match          3.1%; Score 11; DB 2; Length 192;
Best Local Similarity 100.0%; Pred. No. 0.0059;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 324 GSSSSSSSSSS 334
|||||
Db 19 GSSSSSSSSSS 29

RESULT 4
T51989
ethylene responsive element binding factor 2 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T51989
R:Fujimoto, S.Y.; Ohta, M.; Usui, A.; Shinshi, H.; Ohme-Takagi, M.
Plant Cell 12, 393-404, 2000
A:Title: Arabidopsis ethylene responsive element binding factors act as transcriptional activators
A:Reference number: Z25933
A:Accession: T51989
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-243 <FUG>
A:Cross-references: EMBL:AB008104; PIDN:BAA32419.1
C:Genetics:
A:Gene: ERP-2
Query Match          3.1%; Score 11; DB 2; Length 243;
Best Local Similarity 100.0%; Pred. No. 0.0072;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 325 SSSSSSSSSST 335
|||||
Db 193 SSSSSSSSSST 203

RESULT 5
T40941
hypothetical serine-rich protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: T40941
R:Lucas, M.; Gallardin, C.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 1999
A:Reference number: Z21959
A:Accession: T40941
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-262 <LUC>
A:Cross-references: EMBL:AL035259; PIDN:CAA22863.1; GSPDB:GN000068; SPDB:SPCC1322.10
A:Experimental source: strain 972h-; cosmid cl322
C:Genetics:
A:Gene: SPDB:SPCC1322.10
A:Map position: 3
A:Superfamily: serine-rich protein
Query Match          3.1%; Score 11; DB 2; Length 262;
Best Local Similarity 100.0%; Pred. No. 0.0077;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 326 SSSSSSSSSST 336
|||||
Db 156 SSSSSSSSSST 166

RESULT 6
A35419
neutrophil protein - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 18-Jan-1991 #sequence_revision 18-Jan-1991 #text_change 05-Nov-1999
C:Accession: A35419
R:Bellavite, P.; Bazzoni, F.; Cassatella, M.A.; Hunter, K.J.; Bannister, J.V.
Biochem. Biophys. Res. Commun. 170, 915-922, 1990
A:Title: Isolation and characterization of a cDNA clone for a novel serine-rich neutrophil protein
A:Reference number: A35419; MUID:90343818; PMID:1696479
A:Accession: A35419
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-284 <BEL>
A:Cross-references: GB:M55701; NID:gl64672; PIDN:AAA63449.1; PID:gl64673
Query Match          3.1%; Score 11; DB 2; Length 284;
Best Local Similarity 100.0%; Pred. No. 0.0082;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 325 SSSSSSSSSST 335
|||||
Db 145 SSSSSSSSSST 155

RESULT 7
B40722
homeotic protein lin-32 - Caenorhabditis elegans (fragment)
C:Species: Caenorhabditis elegans
C>Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 17-Nov-2000
C:Accession: B40722; A40720
R:Wang, B.B.; Mueller-Immergluck, M.M.; Austin, J.; Robinson, N.T.; Chisholm, A.; Kenyon, C.
Cell 74, 29-42, 1993
A:Title: A homeotic gene cluster patterns the anteroposterior body axis of Caenorhabditis elegans
A:Reference number: B40722; MUID:93327429; PMID:8101474
A:Accession: B40722
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-288 <WAN>
A:Cross-references: GB:L19248
R:Clark, S.G.; Chisholm, A.D.; Horvitz, H.R.
Cell 74, 43-55, 1993
A:Title: Control of cell fates in the central body region of Caenorhabditis elegans by the lin-32 gene
A:Reference number: A40720; MUID:93327430; PMID:8101475
A:Accession: A40720
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 36-288 <CIA>
A:Cross-references: GB:L19639; NID:g304331; PIDN:AAC37168.1; PID:g304332
C:Genetics:
A:Gene: lin-32
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
P:201-257/Domain: homeobox homology <HOX>
Query Match          3.1%; Score 11; DB 2; Length 288;
```

Best Local Similarity 100.0%; Pred. No. 0.0083;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
|||||
DB 59 SSSSSSSSSST 69

RESULT 8
T47832
hypothetical protein T209.60 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000
C:Accession: T47832
R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;
submitted to the Protein Sequence Database, February 2000
A:Reference number: Z24475
A:Accession: T47832
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-306 <NVA>
A:Cross-references: EMBL:AL138658
A:Experimental source: cultivar Columbia; BAC clone T209
C:Genetics:
A:Map position: 3
A:Note: T209.60
C:Superfamily: RING finger homology
F:165-215/Domain: RING finger homology <RRN>

Query Match 3.1%; Score 11; DB 2; Length 306;
Best Local Similarity 100.0%; Pred. No. 0.0088;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GSSSSSSSSSS 334
|||||
DB 123 GSSSSSSSSSS 133

RESULT 9
G96544
hypothetical protein F8A12.3 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G96544
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G96544
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-323 <STO>
A:Cross-references: GB:AE005173; NID:gl1094707; PIDN:AAG29643.1; GSPDB:GN0041
C:Genetics:
A:Gene: F8A12.3
A:Map position: 1

Query Match 3.1%; Score 11; DB 2; Length 323;
Best Local Similarity 100.0%; Pred. No. 0.0093;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GSSSSSSSSSS 334
|||||
DB 165 GSSSSSSSSSS 175

RESULT 10
G83656
hypothetical protein BH0055 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: G83656
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Saeaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: G83656
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <STO>
A:Cross-references: GB:AB001507; GB:BA000004; NID:gl0172612; PIDN:BA03774.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0055

Query Match 3.1%; Score 11; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GSSSSSSSSSS 334
|||||
DB 285 GSSSSSSSSSS 295

RESULT 11
T45525
WSC4 homolog [imported] - yeast (Kluyveromyces marxianus var. lactis)
C:Species: Kluyveromyces marxianus var. lactis, Candida sphaerica
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 31-Jan-2000
C:Accession: T45525
R:Bao, W.G.; Fukuhara, H.
submitted to the EMBL Data Library, July 1999
A:Description: The ubiquitin-encoding genes of Kluyveromyces lactis.
A:Reference number: Z23000
A:Accession: T45525
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-446 <BAO>
A:Cross-references: EMBL:AJ243800; PIDN:CAB50897.1
A:Experimental source: strain 2359/152
C:Genetics:
A:Gene: wsc4

Query Match 3.1%; Score 11; DB 2; Length 446;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
|||||
DB 172 SSSSSSSSSST 182

RESULT 12
E83465
conserved hypothetical protein PA1451 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: E83465
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:110984043
A:Accession: E83465
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-447 <STO>
A:Cross-references: GB:AE004574; GB:AE004091; NID:99947391; PIDN:AAG04840.1; GSPDB:GN001

A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1451

Query Match 3.1%; Score 11; DB 2; Length 447;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 324 GSSSSSSSSSS 334
Db 417 GSSSSSSSSSS 427

RESULT 13

UTR2 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YEL040W
C:Species: Saccharomyces cerevisiae
C:Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 06-Feb-1998
C:Accession: S30839; S50504; S38545
R:Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor, submitted to the EMBL Data Library, February 1993
A:Reference number: S30812
A:Accession: S30839
A:Molecule type: DNA
A:Residues: 1-467 <MUL>
A:Cross-references: GB:U18779; EMBL:L10830; NID:G603625; PID:G603639
R:Dietrich, F.S.
A:Description: The sequence of S. cerevisiae cosmids 8199, 8334, and 9871.
A:Reference number: S50491
A:Accession: S50504
A:Molecule type: DNA
A:Residues: 1-467 <DIR>
A:Cross-references: EMBL:U18779; NID:G603625; PID:G603639; MIPS:YEL040W
R:Melnick, L.; Sherman, F.
J. Mol. Biol. 233, 372-388, 1993
A:Title: The gene clusters ARC and COR on chromosomes 5 and 10, respectively, of Saccharomyces cerevisiae
A:Reference number: S38543; MUID:94016558; PMID:8411151
A:Accession: S38545
A:Molecule type: DNA
A:Residues: 121-129, 'V', 131-290, 'R', 292-353, 'C', 355-467 <MEL>
A:Cross-references: EMBL:S66130; NID:G430829; PID:G430830
C:Genetics:
A:Gene: SGD:UTR2
A:Cross-references: SGD:S0000766; MIPS:YEL040W
A:Map position: 5L

Query Match 3.1%; Score 11; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 325 SSSSSSSSSSS 335
Db 394 SSSSSSSSSSS 404

RESULT 14

S63257
probable membrane protein YNL283c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein N0583
C:Species: Saccharomyces cerevisiae
C:Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 06-Feb-1998
C:Accession: S63257
R:Mesenguy, F.; Dubois, E.; Vierendeels, F.; Scherens, B.; Piertard, A.; Glansdorff, N. submitted to the Protein Sequence Database, April 1996
A:Reference number: S63245
A:Accession: S63257
A:Molecule type: DNA
A:Residues: 1-503 <MES>
A:Cross-references: EMBL:Z71559; NID:g1302356; PID:e239724; PID:g1302357; MIPS:YNL283C
A:Experimental source: strain S288C
C:Genetics:

A:Gene: SGD:WSC2
A:Cross-references: SGD:S0005227; MIPS:YNL283c
A:Map position: 14L
C:Keywords: transmembrane protein
F:330-346/Domain: transmembrane #status predicted <TMM>

Query Match 3.1%; Score 11; DB 2; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 325 SSSSSSSSSSS 335
Db 183 SSSSSSSSSSS 193

RESULT 15

I39411
AF-9 protein - human
C:Species: Homo sapiens (man)
C:Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 29-Aug-1997
C:Accession: I39411
R:Nakamura, T.; Alder, H.; Gu, Y.; Prasad, R.; Canaani, O.; Kamada, N.; Gale, R.P.; Lang Proc. Natl. Acad. Sci. U.S.A. 90, 4631-4635, 1993
A:Title: Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leuk
A:Reference number: A47440; MUID:93281633; PMID:8506309
A:Accession: I39411
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-568 <NAK>
A:Cross-references: GB:L13744; NID:G306448; PID:G306449
C:Genetics:
A:Gene: GDB:MLLT3; AF-9
A:Cross-references: GDB:138172; OMIM:159558
A:Map position: 9p22-9p22

Query Match 3.1%; Score 11; DB 2; Length 568;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 325 SSSSSSSSSSS 335
Db 181 SSSSSSSSSSS 191

Search completed: February 19, 2004, 19:39:16
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 19:34:31 ; Search time 17 Seconds

(without alignments)
995.860 Million cell updates/sec

Title: US-10-041-615-34

Perfect score: 360

Sequence: 1 MIKLGPFYTPTKTKTKKD.....SETPSITQARGSMFLAEHV 360

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 10

Total number of hits satisfying chosen parameters: 112

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	11	3.1	243	1 ERP2_ARATH	O80338 arabidopsis
2	11	3.1	253	1 LI39_CABEL	P34684 caenorhabdi
3	11	3.1	347	1 UTR2_YEAST	P22623 saccharomyc
4	11	3.1	503	1 WSC2_YEAST	P33832 saccharomyc
5	11	3.1	546	1 PHR2_CANAL	O13318 candida alb
6	11	3.1	568	1 AF9_HUMAN	P42568 homo sapien
7	11	3.1	624	1 RN12_HUMAN	Q9NWV2 homo sapien
8	11	3.1	634	1 HWP1_CANAL	P46593 candida alb
9	11	3.1	641	1 FIB2_PETWA	P13573 petromyzon
10	11	3.1	704	1 NPL4_RAT	P41777 rattus norv
11	11	3.1	706	1 CT67_MOUSE	P59114 mus musculu
12	11	3.1	710	1 GSD_ERWCH	Q01565 erwinia chr
13	11	3.1	741	1 NFL1_MOUSE	Q61985 mus musculu
14	11	3.1	772	1 NFL1_HUMAN	Q14494 h nuclear f
15	11	3.1	962	1 GUNA_PSEFL	P10476 pseudomonas
16	11	3.1	1052	1 MGPC_MYCCE	P22747 mycoplasma
17	11	3.1	1237	1 E75A_DROME	P17671 drosophila
18	11	3.1	1341	1 YL78_YEAST	Q05854 saccharomyc
19	11	3.1	1394	1 E75B_DROME	P17672 drosophila
20	11	3.1	1443	1 E75C_DROME	P13055 drosophila
21	11	3.1	1723	1 PM20_CHLPN	Q92812 chlamydia p
22	11	3.1	1823	1 VIT1_CHUN	Q91062 ichtthyomyzo
23	11	3.1	1912	1 VIT1_CHICK	P87498 gallus gall
24	11	3.1	2004	1 CHDB_HUMAN	Q9HCK8 homo sapien
25	10	2.8	98	1 HMEN_HELTR	P23397 helobdella
26	10	2.8	145	1 DH11_GOSHI	P09442 gossypium h
27	10	2.8	170	1 HSB7_HUMAN	Q9UBV9 homo sapien
28	10	2.8	215	1 FGFA_RAT	P70492 rattus norv
29	10	2.8	220	1 SER2_GALME	O96615 gallieria me
30	10	2.8	260	1 ORG1_MOUSE	Q9JKN5 mus musculu
31	10	2.8	304	1 WL13_ARATH	Q9SBV7 arabidopsis
32	10	2.8	334	1 LYTE_BACSU	P54421 bacillus su
33	10	2.8	336	1 CAHC_ARATH	P27140 arabidopsis

ALIGNMENTS

RESULT 1

ERP2_ARATH
ID ERP2_ARATH STANDARD; PRT; 243 AA.
AC O80338
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ethylene responsive element binding factor 2 (ATERF2).
GN ERP2 OR ERP-2 OR AT5G47220 OR MQL5.7.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20181733; PubMed=10715325;
RA Fujimoto S.Y., Ohta M., Usui A., Shinshi H., Ohme-Takagi M.;
RT "Arabidopsis ethylene responsive element binding factors act as
RT transcriptional activators or repressors of GCC box mediated gene
RT expression."
RL Plant Cell 12:393-404 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty p1 and TAC
RL clones."
RL DNA Res 7:31-63 (2000).
CC -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO THE GCC-BOX
CC -!- PATHOGENESIS-RELATED PROMOTER ELEMENT.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Contains 1 AP2/ERF domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC -----
CC ENBL; AB008104; BAA32419.1; -;
CC ENBL; AB018117; BAA97155.1; -;
CC PIR; T51989; T51989.
CC HSSP; O80337; 2GCC.
CC TRANSFAC; T04634; -;
CC InterPro; IPR001471; TF_ERF.
CC Pfam; PF00847; AP2-domain; 1.
CC ProDom; PD001423; TF_ERF; 1.
CC SMART; SM00380; AP2; 1.
CC Transcription regulation; DNA-binding; Nuclear protein.
CC DOMAIN 23 26
CC POLY-GLY.
FT

Q90267 brachydanio
P87179 schizosacch
Q91025 gallus gall
P18417 catharanthu
Q60554 mesocricetu
Q99na9 mus musculu
O35762 rattus norv
P78426 homo sapien
Q03700 rhizopus ni
Q22812 caenorhabdi
P32583 saccharomyc
Q16254 homo sapien

```
FT DNA_BIND 114 179 AP2/BRF.
FT DOMAIN 192 206 POLY-SER.
SQ SEQUENCE 243 AA; 26797 MW; C9A4C9791249B5D4 CRC64;

Query Match
Best Local Similarity 3.1%; Score 11; DB 1; Length 243;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
DB 193 SSSSSSSSSST 203

RESULT 2
LI39 CAEL STANDARD; PRT; 253 AA.
ID LI39 CAEL STANDARD; PRT; 253 AA.
AC F34684; 1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Homeobox protein lin-39.
GN LIN-39 OR CEH-15 OR CO7H6.7.
OS Caenorhabditis elegans.
CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
CC Rhabditidae; Peloderinae; Caenorhabditis.
CC NCBI_TaxID=6239;
RN NCBI_TaxID=6239;
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2.
EX MEDLINE=93327430; PubMed=8101475;
RA Clark S.G., Chisholm A.D., Horvitz H.R.;
RT "Control of cell fates in the central body region of C. elegans by
the homeobox gene lin-39."
RL Cell 74:43-55(1993).
RN NCBI_TaxID=4932;
RP SEQUENCE FROM N.A.
RC STRAIN=B-6441.
EX MEDLINE=94016558; PubMed=8411151;
RA Melnick L., Sherman F.;
RT "The gene clusters ARC and COR on chromosomes 5 and 10, respectively,
of Saccharomyces cerevisiae share a common ancestry."
RL J. Mol. Biol. 233:372-388(1993).
RN NCBI_TaxID=4932;
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
EX PubMed=9169868;
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V."
RL Nature 387:78-81(1997).
CC -!- SIMILARITY: SOME, TO YEAST YGR189C.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L22173; AAA34941.1; -
CC EMBL; S65964; AAD13975.1; -
CC EMBL; S66130; AAB28444.1; -
CC EMBL; U18779; AAB65002.1; ALT_INIT.
CC PIR; S30839; S30839.
CC HSSP; P23904; 1AJO.
CC SGD; S0000766; UTR2.

FT DNA_BIND 114 179 AP2/BRF.
FT DOMAIN 192 206 POLY-SER.
SQ SEQUENCE 243 AA; 26797 MW; C9A4C9791249B5D4 CRC64;

Query Match
Best Local Similarity 3.1%; Score 11; DB 1; Length 253;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
DB 24 SSSSSSSSSST 34

RESULT 3
UTR2 YEAST STANDARD; PRT; 347 AA.
ID UTR2 YEAST STANDARD; PRT; 347 AA.
AC P32623; 1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 30, Last sequence update)
DT 01-OCT-1994 (Rel. 42, Last annotation update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE UTR2 protein (Unknown transcript 2 protein).
GN UTR2 OR YEL040W OR SYGP-ORF18.
OS Saccharomyces cerevisiae (Baker's Yeast).
CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
CC NCBI_TaxID=4932;
RN NCBI_TaxID=4932;
RP SEQUENCE FROM N.A.
RC STRAIN=B-6441.
EX MEDLINE=94016558; PubMed=8411151;
RA Melnick L., Sherman F.;
RT "The gene clusters ARC and COR on chromosomes 5 and 10, respectively,
of Saccharomyces cerevisiae share a common ancestry."
RL J. Mol. Biol. 233:372-388(1993).
RN NCBI_TaxID=4932;
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
EX PubMed=9169868;
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V."
RL Nature 387:78-81(1997).
CC -!- SIMILARITY: SOME, TO YEAST YGR189C.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L22173; AAA34941.1; -
CC EMBL; S65964; AAD13975.1; -
CC EMBL; S66130; AAB28444.1; -
CC EMBL; U18779; AAB65002.1; ALT_INIT.
CC PIR; S30839; S30839.
CC HSSP; P23904; 1AJO.
CC SGD; S0000766; UTR2.
```

DR GO; GO:000277; C:cell wall (sensu Fungi); IDA.
DR GO; GO:000144; C:septin ring (sensu Saccharomycetes); IDA.
DR InterPro; IPR000757; Glyco_hydro_16.
DR Pfam; PF00722; Glyco_hydro_16; 1.
FT DOMAIN 234 322 SER-RICH.
FT DOMAIN 269 283 POLY-SER.
FT CONFLICT 10 10 L -> V (IN REF. 1).
FT CONFLICT 171 171 A -> R (IN REF. 1).
FT CONFLICT 234 234 S -> C (IN REF. 1).
SQ SEQUENCE 347 AA; 36692 MW; 1E8AFB862C4BB328 CRC64;

Query Match 3.1%; Score 11; DB 1; Length 347;
Best Local Similarity 100.0%; Pred. No. 0.0062;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
DB 274 SSSSSSSSSST 284

RESULT 4
WSC2 YEAST STANDARD; PRT; 503 AA.
AC P5832;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cell wall integrity and stress response component 2 precursor.
GN WSC2 OR YNL283C OR N0583.
OS Saccharomycetes cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Messenguy F., Dubois E., Vierendeels F., Scherens B., Pierard A.,
RA Glandsdorff N.,
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 WSC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z71559; CAA96195.1; -
CC PIR; S63257; S63257.
CC SGD; S0005227; WSC2.
CC GO; GO:0005737; C:cytoplasm; IDA.
CC GO; GO:0004888; F:transmembrane receptor activity; IGI.
CC GO; GO:0007047; P:cell wall organization and biogenesis; IGI.
CC GO; GO:0009408; P:response to heat; IGI.
CC GO; GO:0007266; P:rho protein signal transduction; IGI.
CC InterPro; IPR002889; WSC.
CC Pfam; PF01822; WSC; 1.
CC SMART; SMO0321; WSC; 1.
KW Cell wall; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 23
FT CHAIN 24 503
FT CELL WALL INTEGRITY AND STRESS RESPONSE
FT COMPONENT 2.
FT TRANSMEM 326 346
FT DOMAIN 121 319
FT SER/THE-RICH.
FT CARBOHYD 394 394
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 429 429
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 475 475
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 503 AA; 52292 MW; F2392A73C5CBAB50 CRC64;

Query Match 3.1%; Score 11; DB 1; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
DB 183 SSSSSSSSSST 193

RESULT 5
PHR2 CANAL STANDARD; PRT; 546 AA.
AC O13318;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PH responsive protein 2 precursor (pH-regulated protein 2).
GN PHR2.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SC5314;94; PubMed=9315654;
RX MEDLINE=97459694; PubMed=9315654;
RA Muehleischlegel F.A., Fonzi W.A.;
RT PHR2 of Candida albicans encodes a functional homolog of the pH-
RT regulated gene PHR1 with an inverted pattern of pH-dependent
RT expression.";
RL Mol. Cell. Biol. 17:5960-5967(1997).
CC -!- FUNCTION: REQUIRED FOR APICAL CELL GROWTH AND PLAYS AN ESSENTIAL
CC ROLE IN MORPHOGENESIS. MAY BE INTEGRAL TO THE PATHOGENIC ABILITY
CC OF THE ORGANISM (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (potential).
CC -!- INDUCTION: RERESSED AT PH VALUES ABOVE 6 AND PROGRESSIVELY
CC INDUCED AT MORE ACIDIC PH VALUES.
CC -!- SIMILARITY: Belongs to the GAS1 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF011386; AAB80716.1; -
CC InterPro; IPR004886; GAS1.
CC Pfam; PF03198; GAS1; 1.
CC Glycoprotein; Membrane; GPI-anchor; Signal.
FT SIGNAL 1 22
FT CHAIN 23 22
FT PROPEP ? 546
FT DOMAIN 442 523
FT POLY-SER.
FT SER-RICH.
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 40 40
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 59 59
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 149 149
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 410 410
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 546 AA; 58979 MW; 951E0DB9489CEB6 CRC64;

Query Match 3.1%; Score 11; DB 1; Length 546;
Best Local Similarity 100.0%; Pred. No. 0.0091;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 SSSSSSSSSSS 334
DB 496 SSSSSSSSSSS 506

RESULT 6

AF9_HUMAN
ID AF9_HUMAN STANDARD; PRT; 568 AA.
AC P42568;
DT 01-NOV-1995 (Rel. 32, Created)
DE 01-NOV-1995 (Rel. 32, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE AF-9 protein.
GN MLLT3 OR AF9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93281633; PubMed=8506309;
RA Nakamura T., Alder H., Gu Y., Praad R., Canaan O., Kanada N.,
RA Gale R.P., Lange B., Crist W.M., Nowell P.C., Croce C.M., Canaan E.;
RT "Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in
RT acute leukemia share sequence homology and/or common motifs.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:4631-4635(1993).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- DISEASE: Involved in acute leukemias by a chromosomal
CC translocation t(9;11)(p22;q23) that involves MLLT3 and MLL/HRX.
CC The result is a rogue activator protein.
CC -!- SIMILARITY: TO HUMAN MLLT1/ENL.
CC -!- SUBUNIT: TO YEAST TRANSCRIPTION INITIATION FACTOR TFIIF SMALL
CC SUBUNIT (TFG3/ANCL).
CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/AF9.html".
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L13744; AA558361.1; -
CC FIR; I39411; I39411.
CC Genew; HGNC:7136; MLLT3.
CC MIM; 159558; -
CC GO; GO:0005634; C:nucleus; TAS.
CC GO; GO:0007048; P:oncogenesis; TAS.
CC InterPro; IPR005033; YEATS.
CC Pfam; PF03366; YEATS. 1.
KW Transcription regulation; Activator; Nuclear protein;
KW Chromosomal translocation; Proto-oncogene.
FT DOMAIN 149 194 POLY-SER.
FT DOMAIN 295 300 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 383 391 POLY-SER.
FT DOMAIN 466 469 POLY-PRO.
FT SITE 375 375 MLL FUSION POINT (IN ACUTE MYELOID
FT LEUKEMIA PATIENT CO).
FT SITE 481 481 MLL FUSION POINT (IN ACUTE MYELOID
FT LEUKEMIA PATIENT FL).
SQ SEQUENCE 568 AA; 63367 MW; 88020E5B88A13C2 CRC64;
Query Match 3.1%; Score 11; DB 1; Length 568;
Best Local Similarity 100.0%; Pred. No. 0.0095;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 325 SSSSSSSSSSST 335
DB 181 SSSSSSSSSSST 191
RESULT 7
ID_RN12_HUMAN STANDARD; PRT; 624 AA.
AC Q9NWT2; Q9Y598;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RING finger protein 12 (LIM domain interacting RING finger protein)
DE (RING finger LIM domain-binding protein) (R-LIM) (NY-REN-43 antigen).
GN RNFI2 OR RLIM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20469411; PubMed=11013082;
RA Ostendorff H.P., Bossenz M., Mincheva A., Copeland N.G., Gilbert D.J.,
RA Jenkins N.A., Lichter P., Bach I.;
RT "Functional characterization of the gene encoding RLIM, the
RT corepressor of LIM homeodomain transcription factors.";
RL Genomics 69:120-130(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Renal cell carcinoma;
RX MEDLINE=93438124; PubMed=10508479;
RA Scanlan M.J., Gordan J.D., Williamson B., Stockert E., Bander N.H.,
RA Jongeneel C.V., Gure A.O., Jager D., Jager E., Knuth A., Chen Y.-T.,
RA Old L.J.;
RT "Antigens recognized by autologous antibody in patients with renal-
RT cell carcinoma.";
RL Int. J. Cancer 83:456-464(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX Isogai T., Oka T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
RA Watanabe S., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACTS AS A NEGATIVE CO-REGULATOR FOR LIM HOMEODOMAIN
CC TRANSCRIPTION FACTORS. VIA THE RECRUITMENT OF THE SIN3A/HISTONE
CC DEACETYLASE COREPRESSOR COMPLEX.
CC -!- SUBUNIT: ASSOCIATES WITH LIM/HOMEOBOX FACTORS.
CC -!- TISSUE SPECIFICITY: Expressed in many tissues.
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
CC in position 134 and 142.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ271670; CAC14228.1; -
CC EMBL; AF155109; AAD42875.1; ALT_FRAME.
CC EMBL; AK001334; BAA91632.1; -
CC Genew; HGNC:13429; RNFI2.
CC MIM; 300379; -
CC GO; GO:0017053; C:transcriptional repressor complex; NAS.
CC GO; GO:0003714; F:transcription co-repressor activity; NAS.
CC GO; GO:0016481; P:negative regulation of transcription; NAS.
CC InterPro; IPR001841; Znf_ring.
CC Pfam; PF00097; zf-C3HC4; 1.
CC SMART; SM00184; RING; 1.
CC PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
CC PROSITE; PS00518; ZF_RING_2; 1.
CC PROSITE; PS00518; ZF_RING_2; 1.
KW Transcription regulation; Zinc-finger.
FT DOMAIN 422 506 SER-RICH.
FT ZN FING 570 611 RING-TYPE.
FT DOMAIN 453 481 POLY-SER.
FT DOMAIN 500 506 POLY-SER.
FT CONFLICT 126 126 S -> C (IN REF. 1).

```
FT CONFLICT 134 134 N -> D (IN REF. 3).
FT CONFLICT 144 145 YS -> NR (IN REF. 3).
FT CONFLICT 418 418 Y -> H (IN REF. 3).
SQ SEQUENCE 624 AA; 68527 MW; DE3ADE09ACACBCF8 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 0.01; Length 624;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GSSSSSSSSSSS 334
DB 456 GSSSSSSSSSSS 466

RESULT 8
HWPI_CANAL STANDARD; PRT; 634 AA.
AC P46593; O13424; P87019;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Hyphal wall protein 1 (Cell elongation protein 2).
GN HWPI OR ECE2.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SC5314;
RX MEDLINE=96198091; PubMed=8626424;
RA Staab J.F., Ferrer C.A., Sundstrom P.;
RT "Developmental expression of a tandemly repeated, proline-and
RT glutamine-rich amino acid motif on hyphal surfaces on Candida
RT albicans."
RL J. Biol. Chem. 271:6298-6305 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SC5314;
RX Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RA Sharkey L.L., Saporito-Irwin S.M., Fonzi W.A.;
RN [3]
RP SEQUENCE OF 1-225 FROM N.A.
RC STRAIN=SC5314;
RX Staab J.F., Ferrer C.A., Sundstrom P.R.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HYPHAL SURFACE ANTIGEN.
CC -!- SUBCELLULAR LOCATION: HYPHAL SURFACE.
CC -!- TISSUE SPECIFICITY: FOUND IN HYPHAL BUT NOT YEAST FORMS.
CC -!- PFM: MAY BE O-GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
FT REPEAT 132 141
FT REPEAT 142 151
FT REPEAT 152 161
FT REPEAT 162 171
FT REPEAT 172 179
FT REPEAT 180 187
FT DOMAIN 53 58
FT DOMAIN 59 69
FT DOMAIN 65 69
FT DOMAIN 204 207
FT DOMAIN 208 220
FT DOMAIN 291 296
FT DOMAIN 304 307
FT DOMAIN 398 401
FT CONFLICT 141 141
FT CONFLICT 296 296
FT CONFLICT 441 441
FT CONFLICT 492 492
SQ SEQUENCE 634 AA; 65372 MW; F841347576BA6376 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 0.01; Length 634;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 SSSSSSSSSST 335
DB 211 SSSSSSSSSST 221

RESULT 9
FIB2_PETWA STANDARD; PRT; 641 AA.
AC P33573;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Fibrinogen alpha-2 chain precursor [Contains: Fibrinopeptide A].
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92196059; PubMed=1549566;
RA Pan Y., Doolittle R.F.;
RT "cDNA sequence of a second fibrinogen alpha chain in lamprey: an
RT archetypal version alignable with full-length beta and gamma
RT chains."
RL Proc. Natl. Acad. Sci. U.S.A. 89:2066-2070 (1992).
CC -!- FUNCTION: FIBRINOGEN HAS A DOUBLE FUNCTION: YIELDING MONOMERS THAT
CC POLYMERIZE INTO FIBRIN AND ACTING AS A COFACTOR IN PLATELET
CC AGGREGATION.
CC -!- SUBUNIT: HEXAMER CONTAINING 2 SETS OF 3 NONIDENTICAL CHAINS
CC (ALPHA, BETA AND GAMMA), LINKED TO EACH OTHER BY DISULFIDE BONDS.
CC -!- MISCELLANEOUS: CONVERSION OF FIBRINOGEN TO FIBRIN IS TRIGGERED BY
CC THROMBIN, WHICH CLEAVES FIBRINOPEPTIDES A AND B FROM ALPHA & BETA
CC CHAINS, AND THUS EXPOSES THE N-TERMINAL POLYMERIZATION SITES
CC RESPONSIBLE FOR THE FORMATION OF THE SOFT CLOT. THE SOFT CLOT IS
CC CONVERTED INTO THE HARD CLOT BY FACTOR XIIIa WHICH CATALYZES THE
CC EPSILON-(GAMMA-GLUTAMYL)LYSINE CROSS-LINKING BETWEEN GAMMA CHAINS
CC (STRONGER) AND BETWEEN ALPHA CHAINS (WEAKER) OF DIFFERENT
CC MONOMERS.
CC -!- SIMILARITY: Contains 1 fibrinogen C-terminal domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

DR EMBL; M84482; AAA49264.1; -.
DR HSP; P02671; 1FZD.
DR InterPro; IPR002181; Fibrinogen C.
DR Pfam; PF00147; fibrinogen_C; 1.
DR SMART; SW00186; FBG; 1.
DR PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; 1.
KW Blood coagulation; Plasma; Signal; Repeat.
FT SIGNAL 1 23
FT PEPTIDE 24 31
FT CHAIN 32 641
FT SITE 31 32
FT DOMAIN 293 316
FT DISULFID 45 45
FT DISULFID 54 54
FT DISULFID 63 63
FT DISULFID 67 67
FT DISULFID 179 179
FT DISULFID 183 183
FT DISULFID 404 435
FT DISULFID 571 584
FT CARBOHYD 271 271
FT CARBOHYD 397 397
FT CARBOHYD 458 458
SQ SEQUENCE 641 AA; 70756 MW; CA991A8DD698BF55 CRC64;

Query Match 3.1%; Score 11; DB 1; Length 641;
Best Local Similarity 100.0%; Pred. No. 0.01;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 325 SSSSSSSSSST 335
Db 300 SSSSSSSSSST 310

RESULT 10
NP14_RAT
ID NP14_RAT STANDARD; PRT; 704 AA.
AC P4177;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
DE nucleolar phosphoprotein) (Nopp140) (Nucleolar and coiled-body
DE phosphoprotein 1).
GN NOLC1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 292-309 AND 563-601.
RC TISSUE=Liver;
RX MEDLINE=92323542; PubMed=1623516;
RA Meier U.T., Blobel G.;
RT "Nopp140 shuttles on tracks between nucleolus and cytoplasm.";
RL Cell 70:127-138(1992).
RN [2]
RP INTERACTION WITH NOPS AND FIBRILLARIN.
RX MEDLINE=20143579; PubMed=10679015;
RA Yang Y., Isaac C., Wang C., Dragon F., Pogacic V., Meier U.T.;
RT "Conserved composition of mammalian box H/ACA and box C/D small
RT nucleolar ribonucleoprotein particles and their interaction with the
RT common factor Nopp140."
RL Mol. Biol. Cell 11:567-577(2000).
CC -!- FUNCTION: RELATED TO NUCLEOGENESIS. MAY PLAY A ROLE IN THE
CC MAINTENANCE OF THE FUNDAMENTAL STRUCTURE OF THE FIBRILLAR CENTER

```

AND DENSE FIBRILLAR COMPONENT IN THE NUCLEOLUS. IT HAS INTRINSIC GTPASE AND ATPASE ACTIVITIES. MAY PLAY AN IMPORTANT ROLE IN TRANSCRIPTION CATALYZED BY RNA POLYMERASE I (BY SIMILARITY). -!- SUBUNIT: Interacts with Dkci/Nap57, Nop5/Nap65 and fibrillarin. -!- SUBCELLULAR LOCATION: SHUTTLES ON CURVILINEAR TRACKS BETWEEN NUCLEOLUS AND CYTOPLASM. THESE TRACKS EXTEND FROM THE DENSE FIBRILLAR COMPONENT OF THE NUCLEOLUS ACROSS THE NUCLEOPLASM TO A LIMITED NUMBER OF NUCLEAR PORE COMPLEXES. -!- PTM: THIS PROTEIN UNDERGOES RAPID AND MASSIVE PHOSPHORYLATION AND DEPHOSPHORYLATION ON CK-II AND PKC SITES. Nopp140 IS ONE OF THE MOST PHOSPHORYLATED PROTEINS IN THE CELL. -!- SIMILARITY: Contains 1 Lish domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL; M94287; AAA41718.1; -.
EMBL; M94288; AAA41719.1; -.
PIR; B42680; B42680.
InterPro; IPR006594; Lish.
InterPro; IPR003993; treacle.
Pfam; PF05022; SRP40_C; 1.
PRINTS; PR01503; TREACLE.
SMART; SM00667; Lish; 1.
PROSITE; PS00896; Lish; 1.
KW Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding.
FT DOMAIN 10 42
FT DOMAIN 84 570
FT REPEAT 84 95
FT REPEAT 127 138
FT REPEAT 170 181
FT REPEAT 231 242
FT REPEAT 274 285
FT REPEAT 335 346
FT REPEAT 373 384
FT REPEAT 434 445
FT REPEAT 479 490
FT REPEAT 524 535
FT REPEAT 559 570
FT MOD_RES 567 567
FT VARIANT 150 150
FT SEQUENCE 704 AA; 73562 MW; 14DF1BF2DE483EA3 CRC64;
SQ SEQUENCE 704 AA; 73562 MW; 14DF1BF2DE483EA3 CRC64;
Query Match 3.1%; Score 11; DB 1; Length 704;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 324 GSSSSSSSSSS 334
Db 225 GSSSSSSSSSS 235

RESULT 11
CT67_MOUSE
ID CT67_MOUSE STANDARD; PRT; 706 AA.
AC P59114;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protein C20orf67 homolog.
GN C20ORF67.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.


```
CC -1- SUBUNIT: HETERODIMER (PROBABLE). IT MAY FORM A HETERODIMER WITH
CC SEVERAL TRANSCRIPTION FACTORS.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q61985-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q61985-2; Sequence=VSP_000580, VSP_000581;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Widely expressed
CC -1- PTM: COULD BE DEPENDENT ON CKII PHOSPHORYLATION FOR BINDING.
CC -1- SIMILARITY: Belongs to the bZIP family. CNC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X78709; CAA53362.1; -
CC EMBL; AF015881; AAC40108.1; -
CC PIR; I48694; I48694.
CC HSP; P34707; ISKN
CC MGD; MGI:99421; Nfe2l1.
CC InterPro; IPR004827; TF_bZIP.
CC SMART; SM00338; BRLZ; 1.
CC PROSITE; PS50217; BZIP; 1.
CC DNA-binding; Nuclear protein; Phosphorylation; Alternative splicing.
CC DOMAIN 125 259 ASP/GLU-RICH (ACIDIC).
CC DOMAIN 414 447 ASP/GLU-RICH (ACIDIC).
CC DOMAIN 467 486 POLY-SER.
CC DOMAIN 628 643 BASIC MOTIF (BY SIMILARITY).
CC DOMAIN 651 673 LEUCINE-ZIPPER (BY SIMILARITY).
CC DOMAIN 730 737 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC VARSPLIC 1 291 Missing (in isoform Short).
CC VARSPLIC 447 583 /FTId=VSP_000580.
CC Missing (in isoform Short).
CC /FTId=VSP_000581.
CC CONFLICT 318 318 S -> T (IN REF. 2).
CC CONFLICT 387 387 P -> L (IN REF. 2).
CC SEQUENCE 741 AA; 81545 MW; C01E89DD26E7CDEF CRC64;
CC -----
CC Query Match 3.18; Score 11; DB 1; Length 741;
CC Best Local Similarity 100.0%; Pred. No. 0.012;
CC Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 324 GSSSSSSSSSSS 334
CC DQ 466 GSSSSSSSSSSS 476
CC -----
CC RESULT 14
CC ID_NFL1_HUMAN STANDARD; PRT; 772 AA.
CC AC Q14494; Q12877; Q96FN6;
CC DT 15-JUL-1998 (Rel. 36, Created)
CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
CC DE Nuclear factor erythroid 2 related factor 1 (NFE2 related factor 1)
CC (NFE2-related factor 1) (Nuclear factor, erythroid derived 2, like 1)
CC (Transcription factor 1) (Transcription factor HBZ17) (Transcription
CC factor LCR-F1) (Locus control region-factor 1).
CC GN NF2L1 OR NR1 OR TCF11 OR HBZ17.
CC OS Homo sapiens (Human).
CC CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC OX NCBI_TaxID=9606;
CC RN NCB1_TaxID=9606;
CC SEQUENCE FROM N.A. (ISOFORM 1).
```

```
RX MEDLINE=95095252; PubMed=8001966;
RA Luna L., Johnsen O., Skartlien A.H., Pedetour F., Turc-Carel C.,
RA Prydz H., Kolstoe A.-B.;
RT "Molecular cloning of a putative novel human bZIP transcription
RT factor on chromosome 17q22.";
RN Genomics 22:553-562(1994).
RL [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Eye;
RX MEDLINE=23388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.F., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.D., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Mazny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 326-772 FROM N.A.
RX MEDLINE=94310069; PubMed=8036168;
RA Caterina J.J., Donze D., Sun C.W., Ciavatta D.J., Townes T.M.;
RT "Cloning and functional characterization of LCR-F1: a bZIP
RT transcription factor that activates erythroid-specific, human globin
RT gene expression.";
RL Nucleic Acids Res. 22:2383-2391(1994).
CC -1- FUNCTION: ACTIVATES ERYTHROID-SPECIFIC, GLOBIN GENE EXPRESSION.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q14494-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q14494-2; Sequence=VSP_000579;
CC -1- SIMILARITY: Belongs to the bZIP family. CNC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X77366; CAA54555.1; -
CC EMBL; BC010623; AAH10623.1; -
CC EMBL; U08853; AAA20466.1; -
CC PIR; A49672; A49672.
CC PIR; A55004; A55004.
CC HSP; P34707; ISKN.
CC Genew; HGNC:7761; NFE2L1.
CC NM; 163260; -
CC GO; GO:0003712; F:transcription cofactor activity; TAS.
CC GO; GO:0003700; F:transcription factor activity; TAS.
CC GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
CC GO; GO:0006783; P:heme biosynthesis; TAS.
CC GO; GO:0006954; P:inflammatory response; TAS.
CC GO; GO:0006366; P:transcription from Pol II promoter; TAS.
CC InterPro; IPR004827; TF_bZIP.
CC SMART; SM00338; BRLZ; 1.
CC PROSITE; PS50217; BZIP; 1.
```

DR PROSITE; PS00036; BZIP_BASIC; 1.
KW Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW Alternative splicing.
FT DOMAIN 125 288 ASP/GLU-RICH (ACIDIC).
FT ACT_SITE 496 517 POLY-SER.
FT DOMAIN 659 674 BASIC MOTIF.
FT DOMAIN 682 704 LEUCINE-ZIPPER.
FT VARSPIC 242 271 Missing (in isoform 2).
FT /FTID=VSP 000579.
SQ SEQUENCE 772 AA; 84703 MW; C988807C6046BEF5 CRC64;

Query Match 3.1%; Score 11; DB 1; Length 772;
Best Local Similarity 100.0%; Pred.No. 0.012;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GSSSSSSSSSS 334
DB 495 GSSSSSSSSSS 505

RESULT 15

GUNA_PSEFL STANDARD; PRT; 962 AA.
AC P10476;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase) (EGA).
GN CELA.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
CX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sp. Cellulosa;
RX MEDLINE=89127129; PubMed=2851699;
RA Hall J., Gilbert H.J.;
RT "The nucleotide sequence of a carboxymethylcellulase gene from
RL Pseudomonas fluorescens subsp. cellulosa."
Mol. Gen. Genet. 213:112-117(1988).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY E (FAMILY 9 OF GLYCOSYL
HYDROLASES).
CC -1- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X12570; CAA31082.1; -.
HSPF; P07986; IEXG.
DR InterPro; IPR001919; Bac_cellose-bind.
DR InterPro; IPR002883; CBD 5.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR004197; Glyco_hydro_9ig.
DR InterPro; IPR000601; PKD domain.
DR Pfam; PF02013; CEM_10; 1.
DR Pfam; PF00553; CEM_2; 1.
DR Pfam; PF02927; celd_N; 1.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR Pfam; PF00801; PKD; 1.
DR SMART; SM00637; CBD II; 1.
DR SMART; SM00089; PKD; 1.
DR PROSITE; PS00561; CBD BACTERIAL; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.

DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 32 POTENTIAL.
FT CHAIN 33 962 ENDOGLUCANASE A.
FT ACT_SITE 523 523 BY SIMILARITY.
FT ACT_SITE 573 573 BY SIMILARITY.
FT ACT_SITE 582 582 BY SIMILARITY.
FT DOMAIN 608 664 SER-RICH.
FT DOMAIN 823 859 SER-RICH (LINKER).
FT DOMAIN 866 962 CELLULOSE-BINDING (BY SIMILARITY).
FT DISULFID 866 961 BY SIMILARITY.
SQ SEQUENCE 962 AA; 100290 MW; 9E93FE1EAC21A898 CRC64;

Query Match 3.1%; Score 11; DB 1; Length 962;
Best Local Similarity 100.0%; Pred.No. 0.015;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GSSSSSSSSSS 334
DB 607 GSSSSSSSSSS 617

Search completed: February 19, 2004, 19:37:49
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 19, 2004, 19:35:01 ; Search time 40 Seconds
(without alignments)

2322.473 Million cell updates/sec

Title: US-10-041-615-34

Perfect score: 360

Sequence: 1 MIKLGPTPTTKIKTKIKD.....SETPSITQARGSMFLAEHV 360

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 10

Total number of hits satisfying chosen parameters: 894

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL 23:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_mus:*

9: sp_organelle:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	3.6	311	5 Q9V684	Q9V684 drosophila
2	12	3.3	250	5 Q8MN67	Q8MN67 dictyosteli
3	12	3.3	580	5 Q9GUB3	Q9GUB3 dictyosteli
4	12	3.3	624	3 Q8NIV1	Q8NIV1 neurospora
5	12	3.3	967	5 Q9SUI3	Q9SUI3 drosophila
6	12	3.3	1280	5 Q9VWC2	Q9VWC2 drosophila
7	12	3.3	2529	5 Q9VWC0	Q9VWC0 drosophila
8	11	3.1	50	5 Q9NDB3	Q9NDB3 cryptospori
9	11	3.1	59	5 Q9NCR6	Q9NCR6 cryptospori
10	11	3.1	68	5 Q9NCS0	Q9NCS0 cryptospori
11	11	3.1	104	5 Q9NCR8	Q9NCR8 cryptospori
12	11	3.1	107	5 Q9NCR9	Q9NCR9 cryptospori
13	11	3.1	109	5 Q9UZE9	Q9UZE9 caenorhabd
14	11	3.1	131	5 Q9NCS2	Q9NCS2 cryptospori
15	11	3.1	134	16 Q8NQ33	Q8NQ33 corynebacte
16	11	3.1	136	12 O73451	O73451 human papil

17	11	3.1	138	5 Q8NOK1	Q8NOK1 cryptospori
18	11	3.1	141	5 Q8NOJ1	Q8NOJ1 cryptospori
19	11	3.1	145	11 Q8Y58	Q8Y58 mus musculu
20	11	3.1	146	10 Q8VNO	Q8VNO arabidopsis
21	11	3.1	149	11 Q8C5W6	Q8C5W6 mus musculu
22	11	3.1	153	5 Q8NOI2	Q8NOI2 cryptospori
23	11	3.1	156	5 Q8NOI1	Q8NOI1 cryptospori
24	11	3.1	157	5 Q8NOH0	Q8NOH0 cryptospori
25	11	3.1	158	5 Q8NOH1	Q8NOH1 cryptospori
26	11	3.1	158	5 Q8NOG5	Q8NOG5 cryptospori
27	11	3.1	158	5 Q8NOH2	Q8NOH2 cryptospori
28	11	3.1	158	5 Q8NOI0	Q8NOI0 cryptospori
29	11	3.1	159	5 Q8NOG9	Q8NOG9 cryptospori
30	11	3.1	159	5 Q8NOI3	Q8NOI3 cryptospori
31	11	3.1	162	10 Q9MOB4	Q9MOB4 arabidopsis
32	11	3.1	165	5 Q8NOG8	Q8NOG8 cryptospori
33	11	3.1	170	5 Q8NOH9	Q8NOH9 cryptospori
34	11	3.1	170	5 Q8NOH4	Q8NOH4 cryptospori
35	11	3.1	173	5 Q8NOI6	Q8NOI6 cryptospori
36	11	3.1	173	5 Q8NOJ0	Q8NOJ0 cryptospori
37	11	3.1	175	5 Q8NOI7	Q8NOI7 cryptospori
38	11	3.1	178	5 Q8NOK2	Q8NOK2 cryptospori
39	11	3.1	178	5 Q8NOI9	Q8NOI9 cryptospori
40	11	3.1	178	10 Q9LRJ8	Q9LRJ8 arabidopsis
41	11	3.1	178	10 Q8LBR9	Q8LBR9 arabidopsis
42	11	3.1	179	5 Q8NOI4	Q8NOI4 cryptospori
43	11	3.1	181	5 Q8NOI5	Q8NOI5 cryptospori
44	11	3.1	193	10 Q43454	Q43454 glycine max
45	11	3.1	193	11 Q9D6J0	Q9D6J0 mus musculu

ALIGNMENTS

RESULT 1

Q9V684

ID Q9V684

AC Q9V684

DT 01-MAY-2000 (TREMELrel. 13, Created)

DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)

DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)

DE G31169 protein.

GN G31169.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Beeson K.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mactel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AF003823; AAF58547.1; -;
 DR FlyBase; FBgn003704; CG13169.
 SQ SEQUENCE 311 AA; 33572 MW; B75163C3AF5BACE3 CRC64;

Query Match 3.6%; Score 13; DB 5; Length 311;
 Best Local Similarity 100.0%; Pred. No. 0.00015; Indels 0; Gaps 0;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 324 GSSSSSSSSSTP 336
 DB 188 GSSSSSSSSSTP 200

RESULT 2
 Q8NM67 Q8NM67 PRELIMINARY; PRT; 250 AA.
 ID AC Q8NM67
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 CX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4.
 RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
 RA Tungal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
 RT "Sequence and Analysis of Chromosome 2 of *Dictyostelium*.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC117079; AAM43726.1; -;
 DR EMBL; AC117079; AAM43726.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 250 AA; 27881 MW; 163B7C7B3A50264B CRC64;

Query Match 3.3%; Score 12; DB 5; Length 250;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 325 SSSSSSSSSSTP 336
 DB 122 SSSSSSSSSSTP 133

RESULT 3
 Q9GUB3 Q9GUB3 PRELIMINARY; PRT; 580 AA.
 ID AC Q9GUB3
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Myb protein.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 CX NCBI_TaxID=44689;

RN SEQUENCE FROM N.A.
 RP MEDLINE=99265952; PubMed=103131990;
 RX Guo X., Anhard C., Harwood A., Kim H.J., Newell P.C., Gross J.D.;
 RA "A myb-related protein required for culmination in *Dictyostelium*.";
 RL Development 126:2813-2822(1999).
 CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -|- SIMILARITY: CONTAINS 3 MYB-LIKE DOMAINS.
 DR EMBL; AF098507; AAG15274.1; -;
 DR HSP; P06876; IMBG.
 DR InterPro; IPR001005; Myb DNA binding.
 DR Pfam; P00249; myb DNA-binding; 3.
 DR SMART; SM00717; SANT; 3.
 DR PROSITE; PS00037; MYB_1; 1.
 DR PROSITE; PS00030; MYB_3; 3.
 DR DNA-binding; Nuclear protein.
 KW SEQUENCE 580 AA; 66019 MW; F02D9FACB79F3C43 CRC64;

Query Match 3.3%; Score 12; DB 5; Length 580;
 Best Local Similarity 100.0%; Pred. No. 0.0027; Indels 0; Gaps 0;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 324 GSSSSSSSSSTP 335
 DB 191 GSSSSSSSSSTP 202

RESULT 4
 Q8NIV1 Q8NIV1 PRELIMINARY; PRT; 624 AA.
 ID AC Q8NIV1
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN B13H18.080.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 CX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL807374; CAD37056.1; -;
 DR InterPro; IPR000087; Collagen.
 DR Pfam; PF01391; Collagen; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 624 AA; 58999 MW; 73FED5782242E4D9 CRC64;

Query Match 3.3%; Score 12; DB 3; Length 624;
 Best Local Similarity 100.0%; Pred. No. 0.0029; Indels 0; Gaps 0;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 325 SSSSSSSSSSTP 336
 DB 146 SSSSSSSSSSTP 157

RESULT 5
 Q95UI3 Q95UI3 PRELIMINARY; PRT; 967 AA.
 ID AC Q95UI3
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE GH14389p.
 GN CG32529 OR CG11936 OR CG15619.

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY056379; AAL13608.1; -;
 DR FlyBase; FBgn0052529; CG32529.
 DR InterPro; IPR001025; BAH.
 DR Pfam; PF01426; BAH; 1.
 DR SMART; SM00439; BAH; 1.
 SQ SEQUENCE 967 AA, 105914 MW, 7DFD5B4521A3DC2A CRC64;
 Query Match 3.3%; Score 12; DB 5; Length 967;
 Best Local Similarity 100.0%; Pred. No. 0.0043;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 324 GSSSSSSSSSSST 335
 Db 56 GSSSSSSSSSSST 67
 RESULT 6
 Q9VMC2 PRELIMINARY; PRT; 1280 AA.
 ID Q9VMC2
 AC Q9VMC2;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE CG32529 protein.
 GN CG32529 OR CG11936 OR CG15619.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadiet E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Follard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferrari S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuncio J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskaas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp W., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Milburn C., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003513; AAF49024.2; -;
 DR FlyBase; FBgn0052529; CG32529.
 DR InterPro; IPR001025; BAH.
 DR Pfam; PF01426; BAH; 1.
 DR SMART; SM00439; BAH; 1.
 SQ SEQUENCE 1280 AA, 138439 MW, 32331F1FD60E75DB CRC64;
 Query Match 3.3%; Score 12; DB 5; Length 1280;
 Best Local Similarity 100.0%; Pred. No. 0.0054;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 324 GSSSSSSSSSSST 335
 Db 369 GSSSSSSSSSSST 380
 RESULT 7
 Q9VMC0 PRELIMINARY; PRT; 2529 AA.
 ID Q9VMC0
 AC Q9VMC0;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE CG32529 protein.

GN CG32529 OR CG11936 OR CG15619.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V., Chang M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram P.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Minkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sider-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banon J., An H., Baldwin D., Banon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirkas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berwan B., Carlson J.W., Celniker S.E.,
 RA Ciamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003513; AAF49026.2; -;
 DR FlyBase; FBgn0052529; CG32529.
 DR InterPro; IPR001014; Antifreeze_1.
 DR InterPro; IPR001025; BAH.
 DR PRINTS; PR00308; ANTIFREEZE1.
 DR SMART; SMO0439; BAH; 1.
 SQ SEQUENCE 2529 AA; 270675 MW; C93B1726EE7547D4 CRC64;
 Query Match 3.3%; Score 12; DB 5; Length 2529;
 Best Local Similarity 100.0%; Pred. No. 0.0098;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 324 GSSSSSSSSSSST 335
 Db 1618 GSSSSSSSSSSST 1629
 RESULT 8
 Q9NDB3 PRELIMINARY; PRT; 50 AA.
 ID Q9NDB3
 AC Q9NDB3; 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE 15 kDa glycoprotein gp15 (Fragment).
 OS Cryptosporidium parvum.
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
 OC Cryptosporidiidae; Cryptosporidium.
 OX NCBI_TaxID=5807;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GCH1seq; TISSUE=Oocyst;
 RX MEDLINE=20316040; PubMed=10858229;
 RA Strong W.B., Gut J., Nelson R.G.;
 RT "Cloning and sequence analysis of a highly polymorphic Cryptosporidium
 parvum gene encoding a 60-kilodalton glycoprotein and characterization
 of its 15- and 45-kilodalton zote surface antigen products."
 RL Infect. Immun. 68:4117-4134(2000).
 DR EMBL; AF164508; AAF78362.1; -;
 FT NON TER 1 50
 FT NON TER 50 50
 SQ SEQUENCE 50 AA; 4705 MW; D87D5DFE8811C421 CRC64;
 Query Match 3.1%; Score 11; DB 5; Length 50;
 Best Local Similarity 100.0%; Pred. No. 0.0035;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 324 GSSSSSSSSSS 334
 Db 24 GSSSSSSSSSS 34
 RESULT 9
 Q9NCR6 PRELIMINARY; PRT; 59 AA.
 ID Q9NCR6
 AC Q9NCR6; 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Gp15 antigen (Fragment).
 OS Cryptosporidium parvum.
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;

```
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=5807;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HGM07;
RA MEDLINE=20316040; PubMed=10858229;
RX Strong W.B.; Gut J.; Nelson R.G.;
RT "Cloning and sequence analysis of a highly polymorphic Cryptosporidium
parvum gene encoding a 60-kilodalton glycoprotein and characterization
of its 15- and 45-kilodalton zoite surface antigen products.";
RL Infect. Immun. 68:4117-4134(2000).
RT of its 15- and 45-kilodalton zoite surface antigen products.";
RL Infect. Immun. 68:4117-4134(2000).
DR EMBL; AF178693; AAF81991.1; -.
FT NON_TER 1
FT NON_TER 59
SQ SEQUENCE 59 AA; 5742 MW; D9173B91CB482AA3 CRC64;

Query Match 3.1%; Score 11; DB 5; Length 59;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 324 GSSSSSSSSSS 334
Db 29 GSSSSSSSSSS 39

RESULT 10
Q9NCSO PRELIMINARY; PRT; 68 AA.
AC Q9NCSO;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Gp15 antigen (Fragment).
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=5807;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H81;
RA MEDLINE=20316040; PubMed=10858229;
RX Strong W.B.; Gut J.; Nelson R.G.;
RT "Cloning and sequence analysis of a highly polymorphic Cryptosporidium
parvum gene encoding a 60-kilodalton glycoprotein and characterization
of its 15- and 45-kilodalton zoite surface antigen products.";
RL Infect. Immun. 68:4117-4134(2000).
DR EMBL; AF178693; AAF81987.1; -.
FT NON_TER 1
FT NON_TER 68
SQ SEQUENCE 68 AA; 6551 MW; 430C61CA4A22799C CRC64;

Query Match 3.1%; Score 11; DB 5; Length 68;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 324 GSSSSSSSSSS 334
Db 27 GSSSSSSSSSS 37

RESULT 11
Q9NCS8 PRELIMINARY; PRT; 104 AA.
AC Q9NCS8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Gp15 antigen (Fragment).
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=5807;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=HGM07;
RA MEDLINE=20316040; PubMed=10858229;
RX Strong W.B.; Gut J.; Nelson R.G.;
RT "Cloning and sequence analysis of a highly polymorphic Cryptosporidium
parvum gene encoding a 60-kilodalton glycoprotein and characterization
of its 15- and 45-kilodalton zoite surface antigen products.";
RL Infect. Immun. 68:4117-4134(2000).
DR EMBL; AF178695; AAF81989.1; -.
FT NON_TER 1
FT NON_TER 104
SQ SEQUENCE 104 AA; 9926 MW; 3451036523B53C2F CRC64;

Query Match 3.1%; Score 11; DB 5; Length 104;
Best Local Similarity 100.0%; Pred. No. 0.0065;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 324 GSSSSSSSSSS 334
Db 29 GSSSSSSSSSS 39

RESULT 12
Q9NCR9 PRELIMINARY; PRT; 107 AA.
AC Q9NCR9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Gp15 antigen (Fragment).
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=5807;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HS2;
RA MEDLINE=20316040; PubMed=10858229;
RX Strong W.B.; Gut J.; Nelson R.G.;
RT "Cloning and sequence analysis of a highly polymorphic Cryptosporidium
parvum gene encoding a 60-kilodalton glycoprotein and characterization
of its 15- and 45-kilodalton zoite surface antigen products.";
RL Infect. Immun. 68:4117-4134(2000).
DR EMBL; AF178694; AAF81988.1; -.
FT NON_TER 1
FT NON_TER 107
SQ SEQUENCE 107 AA; 10273 MW; 95AACDEF8536673A CRC64;

Query Match 3.1%; Score 11; DB 5; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.0067;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 324 GSSSSSSSSSS 334
Db 29 GSSSSSSSSSS 39

RESULT 13
Q9UZE9 PRELIMINARY; PRT; 109 AA.
AC Q9UZE9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Y46G5A.20 protein.
GN Y46G5A.20.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wallis J.M.;
```

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1999).
DR EMBL; ALJ10485; CAB60362.1; -
DR WormPep; Y4G5A.20; CR242233.
SQ SEQUENCE 109 AA; 11493 MW; 124002FF2443810F CRC64;

Query Match 3.1%; Score 11; DB 5; Length 109;
Best Local Similarity 100.0%; Pred.No. 0.0069;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 325 SSSSSSSSSST 335
Db 40 SSSSSSSSSST 50

RESULT 14

Q9NCS2 PRELIMINARY; PRT; 131 AA.
AC Q9NCS2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Gp15 antigen (Fragment).
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=5807;
[1]
RP SEQUENCE FROM N.A.
RX STRAIN=6HMA1-10;
RX MEDLINE=20316040; PubMed=10858229;
RA Strong W.B.; Gut J.; Nelson R.G.;
RT "Cloning and sequence analysis of a highly polymorphic Cryptosporidium
RT parvum gene encoding a 60-kilodalton glycoprotein and characterization
RT of its 15- and 45-kilodalton zoite surface antigen products."
RL Infect. Immun. 68:4117-4134(2000).
DR EMBL; AF178691; AAF81985.1; -
FT NON_TER 1
FT NON_TER 131
SQ SEQUENCE 131 AA; 12930 MW; 45BE7D7DC31017BA CRC64;

Query Match 3.1%; Score 11; DB 5; Length 131;
Best Local Similarity 100.0%; Pred.No. 0.008;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 324 GSSSSSSSSSS 334
Db 29 GSSSSSSSSSS 39

RESULT 15

Q8N0N3 PRELIMINARY; PRT; 134 AA.
AC Q8N0N3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein Cg11393.
GN Cg11393.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=17118;
[1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;

RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005278; BAB98786.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 134 AA; 13002 MW; DA00A9CCF281E271 CRC64;

Query Match 3.1%; Score 11; DB 16; Length 134;
Best Local Similarity 100.0%; Pred.No. 0.0081;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 325 SSSSSSSSSST 335
Db 42 SSSSSSSSSST 52

Search completed: February 19, 2004, 19:38:43
Job time : 43 secs